

Thiết kế thí nghiệm hoàn toàn ngẫu nhiên

Đỗ Trọng Hợp

Khoa Khoa Học và Kỹ Thuật Thông Tin

Đại Học Công Nghệ Thông Tin TP. Hồ Chí Minh

Example of Completely Randomized Design Experiment

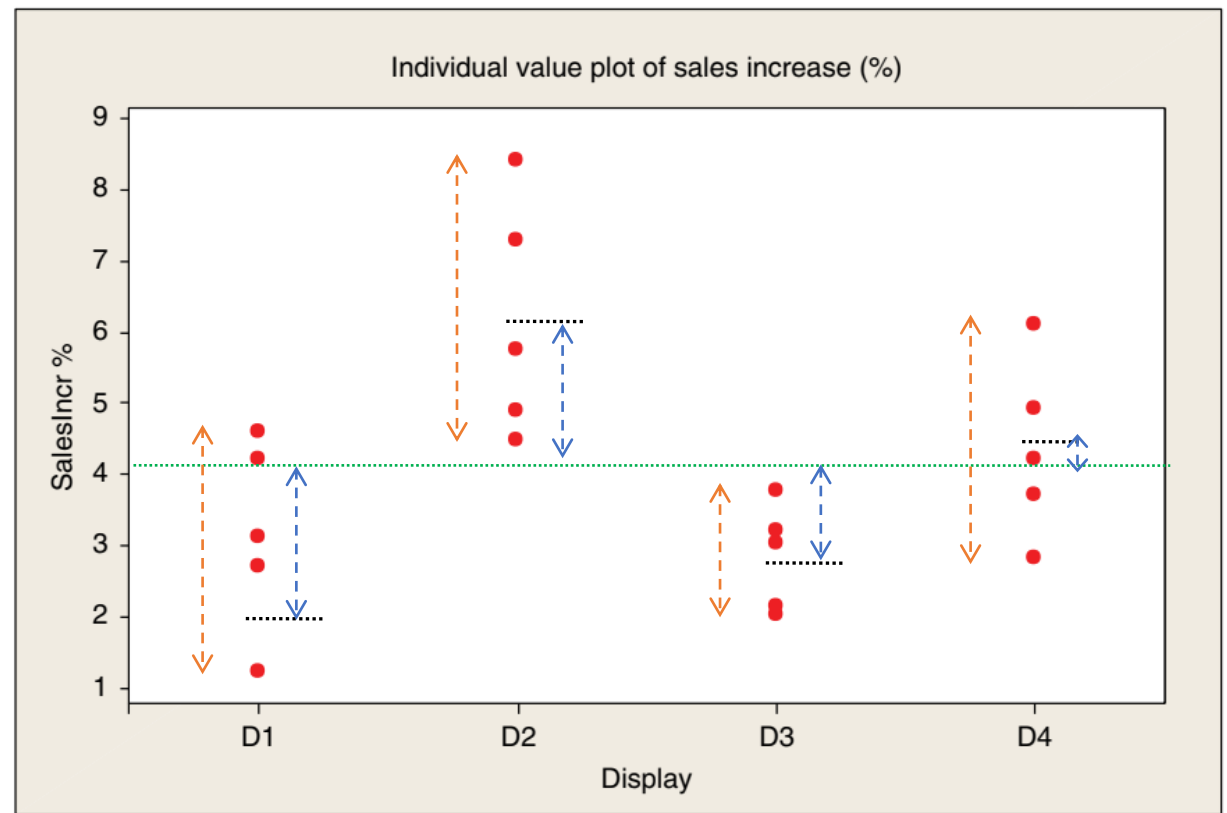
Table 4.1 Sales Data: %Increase in Sales for Four Displays. Each display was installed in five different stores for 1 week.

Sales Increase (%)	Display			
	D1	D2	D3	D4
	4.2	8.4	3.0	4.9
	2.7	4.5	3.8	2.8
	3.1	4.9	2.0	6.1
	4.6	7.3	2.1	4.2
	1.2	5.7	3.2	3.7
Average	3.2	6.2	2.8	4.3
Std. dev.	1.3	1.7	.74	1.2



One-way ANOVA F-test

- Trả lời câu hỏi: có phải 3 (hoặc nhiều hơn) mẫu đến từ những quần thể có cùng mean
- $H_0: \mu_1 = \mu_2 = \mu_3 = \dots \mu_k$
- H_1 : mean các nhóm khác nhau
- $F = \frac{\text{Among group variance}}{\text{Within group variance}}$
- Tính $P(>F)$



ANOVA table

Replication	Treatments				
	1	2	i	k	
1	$y_{1,1}$	$y_{1,2}$...	$y_{1,k}$	
2	$y_{2,1}$	$y_{2,2}$...	$y_{2,k}$	
...	
Total rep	n_1	n_2	n_i	n_k	$n = \sum n_i$
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{i.}$	$\bar{y}_{k.}$	\bar{y}

Source	df	SS	MS=SS/df	F	P-value
Treatments	k-1	$SSA = \sum n_i(\bar{y}_{i.} - \bar{y})^2$	$MSA = SSA/df$	MSA/MSE	$P(>F)$
Error	n-k	$SSE = \sum \sum (y_{ji} - \bar{y}_{i.})^2$	$MSE = SSE/df$		
Total	n-1	$SS = \sum \sum (y_{ji} - \bar{y})^2$			

Tính khoảng tin cậy 95% hai bên

Replication	Displays				
	1	2	3	4	
Total rep	5	5	5	5	20
Group mean	3.2	6.2	2.8	4.3	4.1

Source	df	SS	MS=SS/df	F	P-value
Displays	3	33.91	11.30	6.81	.004
Error	16	26.57	1.66		
Total	19	60.47			

Khoảng tin cậy của μ_1 : $3.2 \pm t_{.025}(16) \sqrt{1.66 / 5} = 3.2 \pm 2.12 \times 1.29 \times 2.24 = (2; 4.4)$

Khoảng tin cậy của μ_2 : $6.2 \pm t_{.025}(16) \sqrt{1.66 / 5} = 6.2 \pm 2.12 \times 1.29 \times 2.24 = (5; 7.4)$

Khoảng tin cậy của $\mu_2 - \mu_1$: $6.2 - 3.2 \pm t_{.025}(16) \sqrt{1.66 \sqrt{2/5}} = 6.2 - 3.2 \pm 2.12 \times 1.29 \times 0.63 = (1.27; 4.72)$

Contrasts

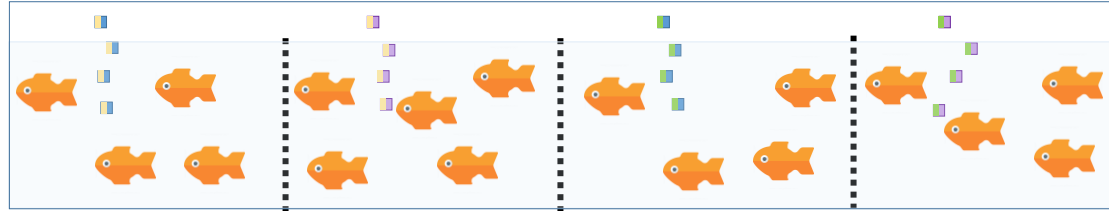
- Định nghĩa:

Tương phản (contrast) là tổ hợp tuyến tính của trung bình các nhóm $\sum w_i \mu_i$, trong đó các hệ số w_i thỏa mãn $\sum w_i = 0$

- Ví dụ
 - So sánh nhóm 1 và 3 trong 4 nhóm: $w=(-1,0,1,0)$ hoặc $w=(1,0,-1,0)$
 - So sánh nhóm 2 và 3 trong 5 nhóm: $w=(0,1,-1,0,0)$ hoặc $w=(0,1,-1,0,0)$
 - So sánh nhóm 1 và nhóm 2,3 trong 3 nhóm: $w=(-1,0.5,0.5)$ hoặc $w=(1,-0.5,-0.5)$
 - So sánh nhóm 1, 2 và nhóm 2,3 trong 4 nhóm: $w=(-0.5,-0.5,0.5,0.5)$ hoặc $w=(0.5,0.5,-0.5,-0.5)$

Ví dụ về tương phản

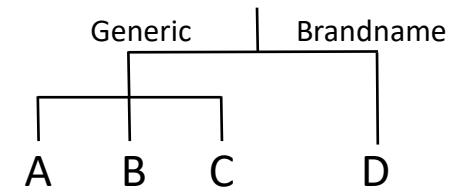
- Thử nghiệm thức ăn cho cá:



- Thức ăn cho cá là kết hợp của protein + carbohydrate
- 4 loại thức ăn là (bột đậu nành + bột ngô), (bột đậu nành + bột mì), (bột cá + bột ngô), (bột cá + bột mì).
- So sánh ảnh hưởng của loại protein thông qua tương phản $w=(-0.5,-0.5,0.5,0.5)$
- So sánh ảnh hưởng của loại carbohydrate qua tương phản $w=(-0.5,0.5,-0.5,0.5)$

- Thử nghiệm loại phân bón

- 4 loại phân bón: A, B, C, D.
- A, B, C là thuộc thương hiệu rẻ, D thuộc thương hiệu nổi tiếng.
- So sánh giữa 2 thương hiệu qua tương phản $w=(1/3,1/3,1/3,-1)$



Suy diễn thống kê cho tương phản

- Ước lượng của tương phản $\sum w_i \mu_i$ là $\sum w_i \bar{y}_i$
- Do variance của \bar{y}_i là σ^2/n_i , ta có

$$\text{Var}\left[\sum w_i \bar{y}_i\right] = \sigma^2 \sum w_i^2 / n_i$$

- Ước lượng của σ^2 là MSE trong bảng ANOVA
- Khoảng tin cậy

$$\sum w_i \bar{y}_i \pm t_{\alpha/2, n-k} \sqrt{MSE} \sqrt{\sum w_i^2 / n_i}$$


SE

Ví dụ

- Thí nghiệm tỉ lệ của gan chuột so với cơ thể sau khi dùng 4 loại thức ăn

i	1	2	3	4
\bar{y}_i	3.75	3.58	3.60	3.92
n_i	7	8	6	8

MSE=.04138

- Thức ăn 1,2,3 làm bởi hãng cũ, thức ăn 4 làm bởi hãng mới. Thí nghiệm muốn so sánh 2 hãng $\rightarrow w=(1/3,1/3,1/3,-1)$

- Ước lượng của tương phản: $\sum w_i \bar{y}_i = \frac{1}{3}3.75 + \frac{1}{3}3.58 + \frac{1}{3}3.60 + (-1)3.92 = -.277$

- Standard error $SE = \sqrt{.04138} \sqrt{\frac{(\frac{1}{3})^2}{7} + \frac{(\frac{1}{3})^2}{8} + \frac{(\frac{1}{3})^2}{6} + \frac{(-1)^2}{8}} = .0847$

- Bậc tự do $df=29-4=25 \rightarrow t_{.025,25}=2.06$

- Khoảng tin cậy 95% của tương phản $-.277 \pm 2.06 \times .0847 = -.277 \pm .174 = (-.451, -.103)$

T-test cho tương phản

- Dùng để kiểm tra giả thuyết $H_0: \sum w_i \mu_i = \delta$
 - Tính $t = \frac{\sum w_i \bar{y}_i - \delta}{SE}$ với $SE = \sqrt{MSE \sum w_i^2 / n_i}$
 - Tính p-value ứng với t tại bậc tự do $df=n-k$
- Ví dụ: ở thí nghiệm thức ăn cho chuột, kiểm tra giả thuyết H_0 : hai hãng khác không khác nhau (kiểm tra 2 đầu)
 - Ta có $\sum w_i \bar{y}_i = -0.277$; $\delta=0$; $SE = 0.847$; $df=25$
 - $t = \frac{-0.277 - 0}{0.847} = -3.27$
 - $P(t \leq -3.27 \mid t \geq 3.27) = 0.003 \rightarrow$ bác bỏ H_0 (công nhận sự khác nhau giữa 2 hãng thức ăn)

F-test cho tương phản

- Tính tổng bình phương của tương phản $SSC = \frac{(\sum w_i y_{bar_i})^2}{\sum w_i^2 / n_i}$
- Bậc tự do của SSC là $df=1 \rightarrow$ Mean square của SSC là $MSC=SSC/1=SSC$
- Dùng F-test để kiểm tra giả thuyết $H_0: \sum w_i \mu_i = 0$
 - $F = MSC/MSE$.
 - Nếu H_0 đúng, F sẽ tuân theo phân phối F với 1 bậc tự do trên tử và $n-k$ bậc tự do dưới mẫu
- Ví dụ: Kiểm định H_0 : không có sự khác nhau giữa 2 hãng thức ăn cho chuột
 - $SSC = \frac{(\frac{1}{3}3.75 + \frac{1}{3}3.58 + \frac{1}{3}3.60 + (-1)3.92)^2}{\frac{(1/3)^2}{7} + \frac{(1/3)^2}{8} + \frac{(1/3)^2}{6} + \frac{(-1)^2}{8}} = \frac{(-.277)^2}{.1733} = .443$
 - $F = MSC/MSE = 0.443/0.04138 = 10.7$
 - Nếu H_0 đúng, F tuân theo phân phối F với bậc tự do 1,25 $\rightarrow P(F \geq 10.7) = 0.003$
 - Nhận xét: p-value của F-test và t-test với $\delta=0$ cho tương phản có giá trị bằng nhau

Tương phản trực giao (orthogonal contrasts)

- Định nghĩa

Two contrasts $\{w\}$ and $\{w^*\}$ are said to be *orthogonal* if

$$\sum_{i=1}^g w_i w_i^* / n_i = 0 \quad .$$

- Đặc tính

- Với k nhóm thì có tối đa k-1 tương phản trực giao lẫn nhau
- Tuy nhiên, có vô hạn cặp tương phản trực giao
- Các tương phản trực giao độc lập với nhau. Do đó, sai số của các tương phản trực giao không tương quan với nhau.
- Tổng các SS của k-1 tương phản trực giao chính là SSA

Ví dụ về tương phản trực giao

- So sánh phác đồ điều trị bệnh
 - 3 phác đồ: control, và 2 phác đồ mới A, B
 - Kích cỡ nhóm là 10, 5, 5 và mean nhóm là 6.3, 6.4, 6.5
 - MSE=0.0225 với 17 bậc tự do
- So sánh giữa control và 2 phác đồ điều trị mới qua tương phản $w(1,-0.5,-0.5)$
- So sánh 2 phác đồ điều trị A và B qua tương phản $w(0,1,-1)$

- 2 tương phản trên trực giao vì

$$\frac{0 \times 1}{10} + \frac{1 \times -0.5}{5} + \frac{-1 \times -0.5}{5} = 0$$

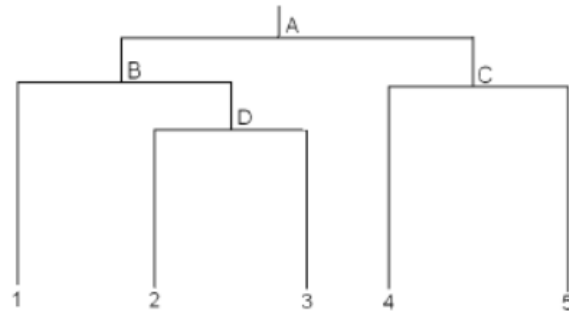
- SS của 2 tương phản trên là:

$$\frac{(6.3 - 0.5 \times 6.4 - 0.5 \times 6.5)^2}{\frac{1}{10} + \frac{(-0.5)^2}{5} + \frac{(-0.5)^2}{5}} = .1125 \qquad \frac{(0 + 6.4 - 6.5)^2}{\frac{0}{10} + \frac{1^2}{5} + \frac{(-1)^2}{5}} = \frac{.01}{.4} = .025$$

- $SSA = 10(6.3 - 6.375)^2 + 5(6.4 - 6.375)^2 + 5(6.5 - 6.375)^2 = .1375$
- Nhận xét: SSC1 lớn gấp 4.5 lần SSC2. Suy ra sự khác nhau giữa phác đồ mới và control cao hơn nhiều so với sự khác nhau giữa 2 phác đồ mới. Nhận xét này không thể có được nếu chỉ dựa vào SSA.

Ví dụ về tương phản trực giao

- Thí nghiệm 5 treatment 1, 2, 3, 4. Mỗi liên quan của một số treatment gần hơn một số khác.
- Các tương phản trực giao được tính toán để phân tích mối liên quan giữa các treatment



Treatments					
Contrasts	1	2	3	4	5
A	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{2}$
B	1	$-\frac{1}{2}$	$-\frac{1}{2}$	0	0
C	0	0	0	1	-1
D	0	1	-1	0	0

- Tương phản A so sánh 3 loại treatment 1,2,3 với 2 loại treatment 4,5. Tương phản B so sánh loại treatment 1 với 2 loại treatment 2,3. Tương phản A và B trực giao nhau.
- Cả 4 tương phản A, B, C, D trực giao với nhau.
- Nhận xét:
 - SSA không đưa ra nhiều thông tin về mối liên quan cụ thể giữa các treatment.
 - Các tương phản không trực giao có thể cho nhiều thông tin hơn nhưng sẽ không đầy đủ, không hệ thống, và còn nhập nhằng giữa các tương phản.
 - Tương phản trực giao giúp phân tích kết quả một cách rõ ràng và hệ thống hơn. Các tương phản độc lập với nhau.

Post hoc analysis

- ANOVA chỉ đưa ra nhận xét về việc có sự khác nhau giữa các nhóm hay không
- Phân tích hậu định (Post hoc analysis) được dùng để so sánh 2 nhóm cụ thể có khác nhau hay không
- Phương pháp đơn giản nhất cho việc này là Least Significant Difference (LSD) test
 - Giả thuyết H_0 : không có sự khác nhau giữa nhóm i và nhóm j
 - Tính $LSD = t_{\alpha/2, df} * s_{ybar_i - ybar_j}$
 - df là bậc tự do của MSE
 - $s_{ybar_i - ybar_j}$ là độ lệch chuẩn của biến ngẫu nhiên $ybar_i - ybar_j$
$$s_{ybar_i - ybar_j} = \sqrt{MSE(1/n_i + 1/n_j)}$$
 - Nếu $|ybar_i - ybar_j| > LSD \rightarrow$ bác bỏ H_0 (nghĩa là công nhận nhóm có khác biệt ở mức ý nghĩa α)
- LSD có nhược điểm lớn là xác suất lỗi false positive cao khi thực hiện multiple testing

Ví dụ LSD-test

Replication	Displays				
	1	2	3	4	
Total rep	5	5	5	5	20
Group mean	3.2	6.2	2.8	4.3	4.1

Source	df	SS	MS=SS/df	F	P-value
Displays	3	33.91	11.30	6.81	.004
Error	16	26.57	1.66		
Total	19	60.47			

- H_0 : không có sự khác biệt giữa display 1 và 2
- $\alpha = 0.05$; $df = 16 \rightarrow t_{.025,16} = 2.12$
- $s_{\bar{y}_1 - \bar{y}_2} = \sqrt{MSE(1/n_1 + 1/n_2)} = \sqrt{1.66 \times (1/5 + 1/5)} = 0.815$
- $LSD = t_{.025,16} * s_{\bar{y}_1 - \bar{y}_2} = 2.12 * 0.815 = 1.73$
- $|\bar{y}_1 - \bar{y}_2| = |3.2 - 6.2| = 3 > 1.73 \rightarrow$ bác bỏ H_0 (công nhận sự khác nhau giữa 2 nhóm 1 và 2)

Hai loại lỗi trong kiểm tra giả thuyết

- **Type I error - false positive** (dương tính giả) là lỗi bác bỏ H_0 khi H_0 đúng
 - Ví dụ: để xét nghiệm 1 loại bệnh, ta so sánh nồng độ thành phần X ở người khám với nồng độ X ở người không bệnh
 - H_0 : nồng độ X của người khám không có khác biệt so với mức bình thường (ở người không bệnh)
 - Nếu phát hiện khác biệt → bác bỏ H_0 (tức là kết luận có bệnh - dương tính)
 - Bác bỏ H_0 (kết luận có khác - dương tính) khi H_0 đúng (thực tế không khác - không có bệnh) → dương tính giả (false positive)
- **Type II error - false negative** (âm tính giả) là lỗi không bác bỏ H_0 khi H_0 sai
 - Ví dụ: để xét nghiệm 1 loại bệnh, ta so sánh nồng độ thành phần X ở người khám với nồng độ X ở người không bệnh
 - H_0 : nồng độ X của người khám không có khác biệt so với mức bình thường (ở người không bệnh)
 - Nếu phát hiện khác biệt → bác bỏ H_0 (tức là kết luận có bệnh - dương tính)
 - Nếu không phát hiện khác biệt → không bác bỏ H_0 (tức là kết luận không bệnh - âm tính)
 - Không bác bỏ H_0 (kết luận không khác - âm tính) khi H_0 sai (thực tế khác - có bệnh) → âm tính giả (false negative)

Vấn đề của post hoc analysis - multiple comparison

- Multiple comparison xảy ra khi kiểm tra giả thuyết H_0 mà H_0 bao gồm nhiều giả thuyết H_{0i} riêng lẻ. Ta bác bỏ H_0 khi bác bỏ tất cả H_{0i}
 - Ví dụ: để xét nghiệm 1 loại bệnh, ta so sánh nồng độ n thành phần X_i ở người khám với mức bình thường ở người không bệnh.
 - Nếu chỉ số của tất cả n thành phần X_i đều bình thường, ta kết luận không bệnh
 - H_0 : không có sự khác biệt giữa người khám và người bình thường (tức là không bệnh) bao gồm n giả thuyết riêng lẻ H_{0i}
 - H_{0i} : không có khác biệt giữa thành phần X_i ở người khám và người bình thường
 - Bác bỏ ít nhất 1 giả thuyết H_{0i} (nghĩa là phát hiện sự bất thường ở thành phần X_i) \rightarrow bác bỏ H_0 (tức là kết luận có bệnh)
 - Xác suất false positive của H_0 (bác bỏ H_0 khi H_0 đúng) là xác suất xảy ra ít nhất 1 lỗi false positive trong n test H_{0i}
- Vấn đề multiple comparison
 - Giả sử H_0 gồm 20 test riêng lẻ H_{0i} và mức ý nghĩa của mỗi test riêng lẻ là 0.05
 - Xác suất xuất hiện lỗi false positive ở mỗi test H_{0i} là 0.05
 - Xác suất không xuất hiện lỗi false positive ở mỗi test H_{0i} là $1-0.05$
 - Xác suất không xuất hiện lỗi false positive ở cả 20 test là $(1-0.05)^{20}$
 - Xác suất lỗi false positive của H_0 = xác suất xuất hiện ít nhất 1 lỗi false positive ở 20 test là $1-(1-0.05)^{20}=64\%$
- Do xác suất gặp lỗi false positive cao, LSD chỉ được dùng khi
 - Kết quả F-test cho thấy có sự khác nhau giữa các nhóm
 - test 2 nhóm mà ta xác định sẽ test khi thực hiện thí nghiệm

Bonferroni Correction

- Bonferroni correction dùng để giảm xác suất lỗi false positive khi thực hiện multiple comparison
 - Giả sử ta cần thực hiện n test và mức ý nghĩa mong muốn là α
 - Bonferroni correction đơn giản sẽ điều chỉnh lại mức ý nghĩa: $\alpha' = \alpha/n$
 - Mức ý nghĩa mới α' sẽ được dùng cho n test
- Ví dụ:
 - Thực hiện 20 test với mức ý nghĩa 0.05
 - Qua Bonferroni correction, mức ý nghĩa được điều chỉnh thành $0.05/20 = 0.0025$
 - Xác suất lỗi false positive ở ít nhất 1 test trong 20 test là $P = 1 - (1 - 0.0025)^{20} = 0.0488 < 0.05$
- Vấn đề
 - Với n lớn, mức ý nghĩa điều chỉnh trở nên quá nhỏ \rightarrow khó để bác bỏ H_0 cho dù H_0 thật sự sai
 - Khi n lớn, α' quá nhỏ \rightarrow xác suất không bác bỏ H_0 khi H_0 thật sự sai (false negative - âm tính giả) trở nên rất cao
 - Ta có thể nói phương pháp Bonferroni rất bảo thủ. Với 1 ca dương tính thì khả năng rất cao là dương tính thật sự, nhưng với 1 ca âm tính thì không chắc có thật sự âm tính hay không
 - Bonferroni chỉ được dùng khi số lượng test không nhiều. Hoặc được dùng khi ta chỉ cần tìm một vài ca dương tính và yêu cầu độ chính xác cao cho các ca dương tính đấy.

Scheffe F-test for multiple contrasts

- Nhắc lại F-test cho một contrast
 - Tính $F = \frac{MSC}{MSE} = \frac{SSC}{MSE}$
 - Tính $P(>F)$ theo phân phối F với 1 bậc tự do trên tử và $n-k$ bậc tự do dưới mẫu. Nếu $P(>F) < \alpha \rightarrow$ bác bỏ H_0
- Scheffe F-test cho nhiều contrast
 - Tính $F = \frac{SSC/(k-1)}{MSE}$
 - Tính $P(>F)$ theo phân phối F với $k-1$ bậc tự do trên tử và $n-k$ bậc tự do dưới mẫu. Nếu $P(>F) < \alpha \rightarrow$ bác bỏ H_0
 - Lưu ý: Scheffe test kiểm tra đuôi bên phải
- Nhận xét:
 - Ở Scheffe F-test, giá trị F tính được thấp hơn \rightarrow p-value sẽ cao hơn \rightarrow khó bác bỏ H_0 hơn
 - Scheffe F-test giảm xác suất false positive (bác bỏ H_0 khi H_0 đúng)
 - Xác suất false negative của Scheffe F-test sẽ tăng khi số lượng nhóm tăng
 - Thực tế Scheffe F-test thường bảo thủ hơn cả Bonferroni (tức là xác suất false negative cao hơn)

Scheffe T-test for multiple contrasts

- Nhắc lại t-test cho một contrast

- Kiểm tra $\frac{\sum w_i ybar_i - \delta}{SE} > t_{\alpha, n-k}$ với $SE = \sqrt{MSE \sum w_i^2 / n_i}$

- Scheffe t-test cho nhiều contrast

- Kiểm tra $\frac{\sum w_i ybar_i - \delta}{SE} > \sqrt{(k-1)F_{\alpha, k-1, n-k}}$

- Lưu ý: Scheffe test là kiểm tra đuôi phải

- Tính khoảng tin cậy của contrast theo Scheffe t-test

$$\sum w_i ybar_i \pm \sqrt{(k-1)F_{\alpha, k-1, n-k}} \times SE$$

Ví dụ

- Thí nghiệm ảnh hưởng của mưa axit đến mầm cây bạch dương

pH	4.7	4.0	3.3	3.0	2.3
weight	.337	.296	.320	.298	.177
n	48	48	48	48	48

- MSE = 0.0119 với bậc tự do 235

- Từ dữ liệu ta thấy trọng lượng mầm cây tại mức pH 2.3 thấp hơn hẳn so với các mức pH từ 3.0 đến 4.7

- Ta tính tương phản $w(1/4, 1/4, 1/4, 1/4, -1)$. Giá trị của tương phản này là

$$\sum w_i ybar_i = \frac{.337 + .296 + .320 + .298}{4} - .177 = .1357 \quad SE = \sqrt{.0119 \left(\frac{.0625}{48} + \frac{.0625}{48} + \frac{.0625}{48} + \frac{.0625}{48} + \frac{1}{48} \right)} = .0176 .$$

- Ta có `sqrt(4*qf(0.01,4,235,lower.tail = FALSE)) = 3.688`

- Khoảng tin cậy 99% là $0.1357 \pm \sqrt{4F_{.01,4,235}} \times 0.0176 = (0.07, 0.2) \rightarrow$ kết luận có sự khác biệt tại mức pH 2.3

- Hoặc ta có thể thực hiện Scheffe F-test $SSC = \frac{(\sum w_i ybar_i)^2}{\sum w_i^2 / n_i} = \frac{0.1357^2}{0.07291} = 0.2525 \quad F = \frac{SSC / (k-1)}{MSE} = \frac{0.2525 / 4}{0.0119} = 5.3$

$P(F > 5.3) = \text{pf}(5.3, 4, 235, \text{lower.tail} = \text{FALSE}) = 0.0004 < 0.01 \rightarrow$ kết luận có sự khác biệt tại mức pH 2.3

- Hoặc ta có thể thực hiện Scheffe t-test

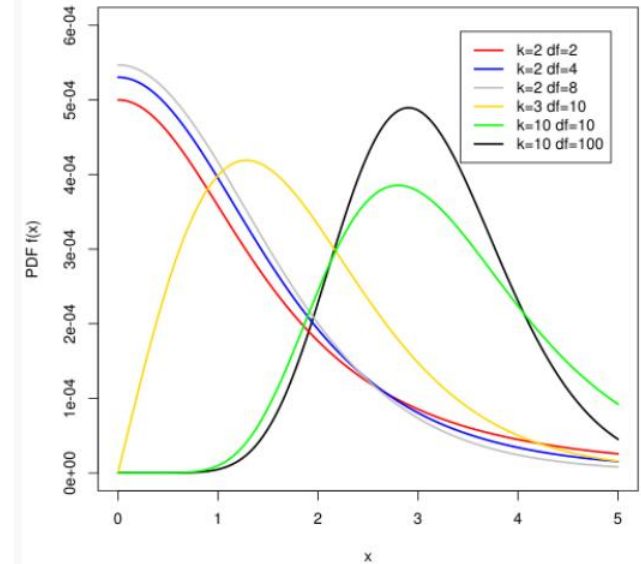
$$\frac{|\sum w_i ybar_i - 0|}{SE} = \frac{0.1357}{0.0176} = 7.71 > \sqrt{4F_{.01,4,235}} = 3.688 \rightarrow \text{kết luận có sự khác biệt tại mức pH 2.3}$$

Studentized range distribution

- Giả sử ta lấy k mẫu kích thước n từ k quần thể có cùng normal distribution $N(\mu, \sigma^2)$
- \bar{y}_{\max} và \bar{y}_{\min} là max và min của k trung bình của k mẫu
- Phương sai gộp của tất cả k mẫu là MSE
- Biến ngẫu nhiên dưới đây tuân theo Studentized range distribution

$$q = \frac{\bar{y}_{\max} - \bar{y}_{\min}}{\sqrt{MSE} / \sqrt{n}}$$

- Lưu ý: $n_1=n_2=\dots=n_k=n$ là kích thước mỗi mẫu. Kích thước của toàn bộ k mẫu là $k*n$
- Studentized range distribution phụ thuộc vào 2 biến là k (số nhóm) và $df=(n-1)*k$ là bậc tự do của MSE
- Nếu H_0 đúng (các nhóm có cùng mean) thì xác suất để gặp giá trị q lớn sẽ nhỏ
- Nếu H_0 sai (các nhóm có mean khác nhau) thì xác suất để gặp giá trị q lớn sẽ lớn
- Suy ra ta có thể dùng Studentized range distribution để bác bỏ hoặc chấp nhận H_0



Tukey Honest Significant Difference (HSD)

- Dùng để so sánh trung bình \bar{y}_i và \bar{y}_j của 2 nhóm i, j
- Tính $q = \frac{|\bar{y}_i - \bar{y}_j| \times \sqrt{2}}{\sqrt{MSE} * \sqrt{1/n_i + 1/n_j}}$
- Tính ngưỡng $q_{\alpha, k, n*k-k}$ trong đó k là số nhóm, n là số phần tử mỗi nhóm, $n*k$ là tổng số phần tử
- Nếu $q > q_{\alpha, k, n*k-k} \rightarrow$ bác bỏ H_0 (2 nhóm khác nhau)
- Hoặc tương tự, tính $p\text{-value} = P(>q)$, nếu $p\text{-value} < \alpha \rightarrow$ bác bỏ H_0
- Khoảng tin cậy $\bar{y}_i - \bar{y}_j \pm \frac{q_{\alpha, k, (n-1)*k}}{\sqrt{2}} \sqrt{MSE} \sqrt{1/n + 1/n}$
- Lưu ý: Tukey HSD kiểm tra đuôi bên phải

Ví dụ

- Thí nghiệm ảnh hưởng của vi khuẩn lên lượng amino axit tự do trong phô mai
- Có 4 nhóm, mỗi nhóm 2 phần tử (có tổng cộng 8 miếng phô mai được làm)
- Ta có $n=2$, $k=4$, $MSE = 0.1572$
- Giả sử $\alpha=0.1$
- $q_{0.1,4,4} = \text{qtukey}(0.1,4,4,\text{lower.tail} = \text{FALSE}) = 4.586$

None	Strain added		
	A	B	A&B
4.195	4.125	4.865	6.155
4.175	4.735	5.745	6.488

mean: 4.19 4.43 5.31 6.32

- Kiểm tra H_0 : không có sự khác nhau giữa nhóm 1 và nhóm 4 với mức ý nghĩa 10%

$$q = \frac{|ybar_4 - ybar_1| \sqrt{2}}{\sqrt{MSE} * \sqrt{1/2 + 1/2}} = \frac{(6.32 - 4.19) \times \sqrt{2}}{\sqrt{0.1572} * \sqrt{1}} = 7.597$$

- $q > q_{0.1,4,4} \rightarrow$ bác bỏ H_0
- $P(>q) = P(>7.59) = \text{ptukey}(7.597,4,4,\text{lower.tail} = \text{FALSE}) = 0.019 \rightarrow$ bác bỏ H_0
- Khoảng tin cậy 90%

$$ybar_4 - ybar_1 \pm \frac{q_{\alpha,k,(n-1)*k}}{\sqrt{2}} \sqrt{MSE} \sqrt{1/2 + 1/2} = 6.32 - 4.19 \pm 3.243 * \sqrt{0.1572} = (0.8, 3.4)$$

Tukey test in R

```
> T1=c(4.195,4.175)
> T2=c(4.125,4.735)
> T3=c(4.865,5.745)
> T4=c(6.155,6.488)
> x=c(T1,T2,T3,T4)

> Treatments=c(rep("T1",2),rep("T2",2),rep("T3",2),rep("T4",2))
> data=data.frame(x,Treatments)
> av = aov(x~Treatments)
```

```
> summary(av)
              Df Sum Sq Mean Sq F value
Treatments    3  5.628   1.8760   11.93
Residuals     4  0.629   0.1572
Pr(>F)
Treatments 0.0183 *
```

```
> TukeyHSD(av,conf.level = 0.9)
Tukey multiple comparisons of means
90% family-wise confidence level
```

```
Fit: aov(formula = x ~ Treatments)
```

		diff	lwr	upr	p adj
T2-T1	0.2450	-1.0408889	1.530889	0.9212486	
T3-T1	1.1200	-0.1658889	2.405889	0.1462787	
T4-T1	2.1365	0.8506111	3.422389	0.0194380	
T3-T2	0.8750	-0.4108889	2.160889	0.2636576	
T4-T2	1.8915	0.6056111	3.177389	0.0296048	
T4-T3	1.0165	-0.2693889	2.302389	0.1870664	

Do khoảng tin cậy T4-T1 và T4-T2 không chứa 0 → chỉ có sự khác biệt giữa 2 cặp nhóm này là có ý nghĩa thống kê

So sánh LSD, Bonferroni, Scheffe, Tukey HSD

```
> pairwise.t.test(x,Treatments,p.adjust.method = "none",alternative = "greater")
```

Pairwise comparisons using t tests with pooled SD

data: x and Treatments

	T1	T2	T3
T2	0.2850	-	-
T3	0.0238	0.0460	-
T4	0.0029	0.0044	0.0312

P value adjustment method: none

```
> pairwise.t.test(x,Treatments,p.adjust.method = "bonferroni",alternative = "greater")
```

Pairwise comparisons using t tests with pooled SD

data: x and Treatments

	T1	T2	T3
T2	1.000	-	-
T3	0.143	0.276	-
T4	0.017	0.027	0.187

P value adjustment method: bonferroni

```
> library(DescTools)
```

```
> ScheffeTest(av, which = NULL, contrasts = NULL, conf.level = 0.9)
```

Posthoc multiple comparisons of means : Scheffe Test
90% family-wise confidence level

\$Treatments	diff	lwr.ci	upr.ci	pval
T2-T1	0.2450	-1.1609541	1.650954	0.9391
T3-T1	1.1200	-0.2859541	2.525954	0.1841
T4-T1	2.1365	0.7305459	3.542454	0.0264 *
T3-T2	0.8750	-0.5309541	2.280954	0.3180
T4-T2	1.8915	0.4855459	3.297454	0.0398 *
T4-T3	1.0165	-0.3894541	2.422454	0.2317

```
> TukeyHSD(av,conf.level = 0.9)
```

Tukey multiple comparisons of means
90% family-wise confidence level

Fit: aov(formula = x ~ Treatments)

\$Treatments	diff	lwr	upr	p adj
T2-T1	0.2450	-1.0408889	1.530889	0.9212486
T3-T1	1.1200	-0.1658889	2.405889	0.1462787
T4-T1	2.1365	0.8506111	3.422389	0.0194380
T3-T2	0.8750	-0.4108889	2.160889	0.2636576
T4-T2	1.8915	0.6056111	3.177389	0.0296048
T4-T3	1.0165	-0.2693889	2.302389	0.1870664