

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

CRD with one factor

Completely Randomized Design experiment with one factor

Thuật ngữ và khái niệm

- **Yếu tố (Factor)**
 - Là biến số độc lập được thiết lập bởi thí nghiệm.
 - Có thể là biến số định lượng hoặc định tính, gián đoạn hoặc liên tục.
- **Kết quả (Response)**
 - Là kết quả thu đo được của thí nghiệm
 - Có thể gọi là biến số phụ thuộc (vào các yếu tố được thiết lập) của thí nghiệm
- **Yếu tố chính (Primary factor)**
 - Là yếu tố cần tìm hiểu ảnh hưởng của nó lên kết quả
- Ví dụ:
 - Thí nghiệm tác dụng của 2 loại phân bón A và B lên sản lượng lúa
 - Có 1 yếu tố chính là phân bón.
 - So sánh 3 loại phân bón A, B, C và nhiệt độ lên sản lượng lúa
 - Có 2 yếu tố chính là (1) phân bón và (2) nhiệt độ
 - So sánh 3 loại phân bón A, B, C lên sản lượng lúa. Nhiệt độ trong thí nghiệm được giữ ở mức cố định.
 - Có 1 yếu tố chính là phân bón. Nhiệt độ là yếu tố cố định (constant factor).

Thuật ngữ và khái niệm

- Mức (Level)
 - Là loại hình hoặc giá trị của 1 yếu tố.
- Ví dụ:
 - Yếu tố phân bón: 2 level gồm loại A và loại B
 - Yếu tố nhiệt độ: vô số level là các giá trị liên tục của nhiệt độ
 - Yếu tố mật độ trồng lúa: 3 level gồm 10, 20, 40 khóm/m²

Thuật ngữ và khái niệm

- **Nghiệm thức (Treatment)**
 - Là những mức của một yếu tố chính được thử nghiệm để so sánh.
- **Ví dụ**
 - So sánh tác dụng thuốc. Mỗi nhóm bệnh nhân được thử (treated) 1 loại thuốc khác nhau. Mỗi loại thuốc là một nghiệm thức (treatment)
 - Thí nghiệm ảnh hưởng của tốc độ lên độ mòn bánh xe. Mỗi xe được thử 1 tốc độ. Mỗi tốc độ là 1 nghiệm thức.
 - So sánh năng suất giống lúa. Mỗi ô ruộng được trồng 1 giống lúa. Mỗi giống lúa là một nghiệm thức.

Thuật ngữ và khái niệm

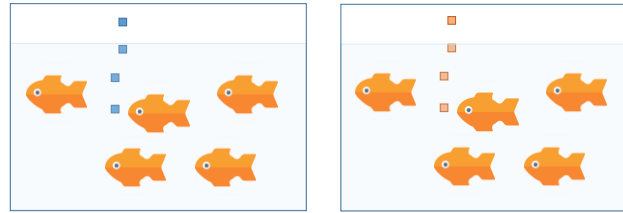
- Đơn vị thí nghiệm (Experiment unit)
 - Là đối tượng nhận các treatment.
 - Mỗi đối tượng có thể nhận 1 treatment khác nhau.
- Ví dụ
 - Thí nghiệm thử thuốc. Mỗi bệnh nhân thử thuốc là 1 E unit.
 - So sánh phương pháp dạy. Có 10 lớp, 5 lớp thử pp offline, 5 lớp thử pp online. Mỗi lớp là 1 đvtn.
 - So sánh 2 loại vật liệu làm giày. Có 10 người, mỗi người thử 1 đôi giày làm từ 1 loại vật liệu. Mỗi người là 1 đvtn.
 - So sánh 2 loại vật liệu làm giày. Có 10 người, mỗi người mang 2 chiếc giày làm từ 2 loại vật liệu. Mỗi chân là 1 đvtn.

CRD with one factor

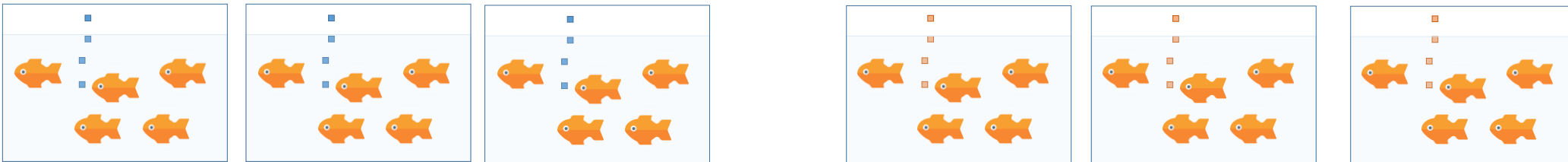
- Thí nghiệm kiểm tra tác dụng của 1 yếu tố chính
- Yếu tố chính có từ 2 mức trở lên
- Mỗi treatment phải được áp dụng cho nhiều experiment unit (nguyên tắc replication)
- Số lượng experiment unit cho mỗi treatment có thể khác nhau
- Các treatment được áp dụng cho các experiment unit theo cách ngẫu nhiên (nguyên tắc randomization)

Replication

- Thí nghiệm 2 loại thức ăn cho cá. Có 2 ao cá, mỗi ao được thử một loại thức ăn. Trọng lượng cá trung bình mỗi ao được đo để so sánh.



- Vấn đề: mỗi ao có thể có môi trường khác nhau nên cá ở hai ao có thể phát triển khác nhau. Do đó không thể kết luận về tác dụng của thức ăn
- Replication principle:** the experiment should be repeated more than once. Thus, each treatment is applied in **many experimental units** instead of one. By doing so the statistical accuracy of the experiments is increased.



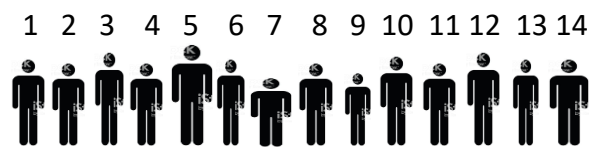
- Để đáp ứng nguyên tắc replication, mỗi loại thức ăn (treatment) sẽ được thử với nhiều ao (experiment unit).

Randomization

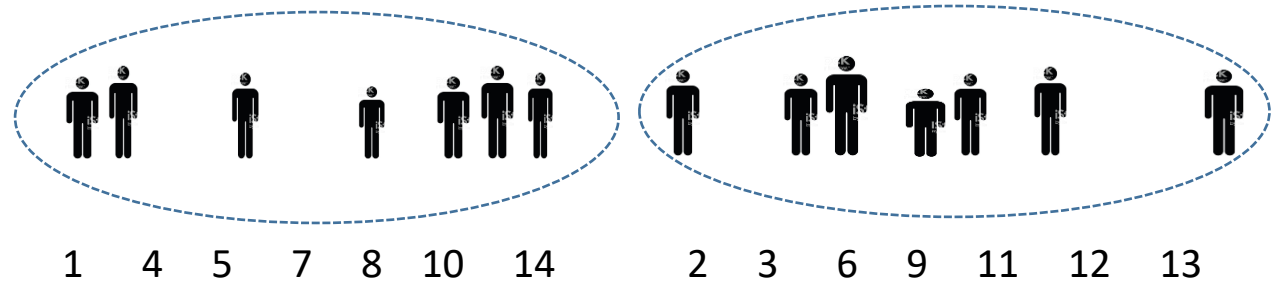
- So sánh tác dụng 2 loại thuốc. Mỗi ứng viên (experiment unit) sẽ được áp dụng 1 loại thuốc (treatment)



- Vấn đề: mỗi ứng viên có thể trạng khác nhau, các loại thuốc được áp dụng thế nào lên các ứng viên để tránh thiên vị (bias)
- Nguyên tắc randomization:** các treatments được áp dụng cho các Experiment unit một cách hoàn toàn ngẫu nhiên. Cơ hội mà mỗi E unit nhận 1 treatment bất kỳ là như nhau.



Randomize



- Đánh số cho mỗi ứng viên. Dùng máy tính chọn ngẫu nhiên 7 số cho mỗi loại thuốc

Ngẫu nhiên (randomized) \neq bừa bãi (haphazard)

- Giả sử chúng ta cần áp dụng 4 treatments A,B,C,D cho 16 Experiment unit
 - Cách thực hiện
 - Đánh số thứ tự 16 người. Dùng máy tính sinh ngẫu nhiên 4 số cho mỗi nhóm.
 - Đánh dấu 16 lá thăm với A,B,C,D. Mỗi E unit bốc ngẫu nhiên 1 lá để chọn loại treatment.

} Randomized

 - 4 người bất kì gặp đầu tiên sẽ áp dụng loại A, và tiếp tục cho B,C,D.
 - Với mỗi E unit, áp dụng treatment A,B,C,D tùy theo kim giây nằm ở phần tư thứ 1,2,3,4 của đồng hồ
- } Haphazard
- Bừa bãi (haphazard) tuy cho kết quả không dự đoán được (predictable) và không tất định (deterministic) nhưng không phải ngẫu nhiên. Có nhiều cơ chế ảnh hưởng đến kết quả mà ta không biết được, gây khó khăn cho việc thiết lập lại thí nghiệm.

Lý do cần ngẫu nhiên

- Tránh những thiên vị (bias) làm ảnh hưởng đến kết quả thí nghiệm
 - Trong thí nghiệm thuốc, mỗi đối tượng có tình trạng khác nhau (giới tính, cân nặng, độ tuổi, sức khỏe,...)
 - Trong thí nghiệm phân bón, mỗi cây cà chua có nhiều đặc điểm (gene của cây, vị trí trồng, tình trạng sâu bệnh,...)

Ví dụ thí nghiệm trong marketing

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



Phân tích kết quả

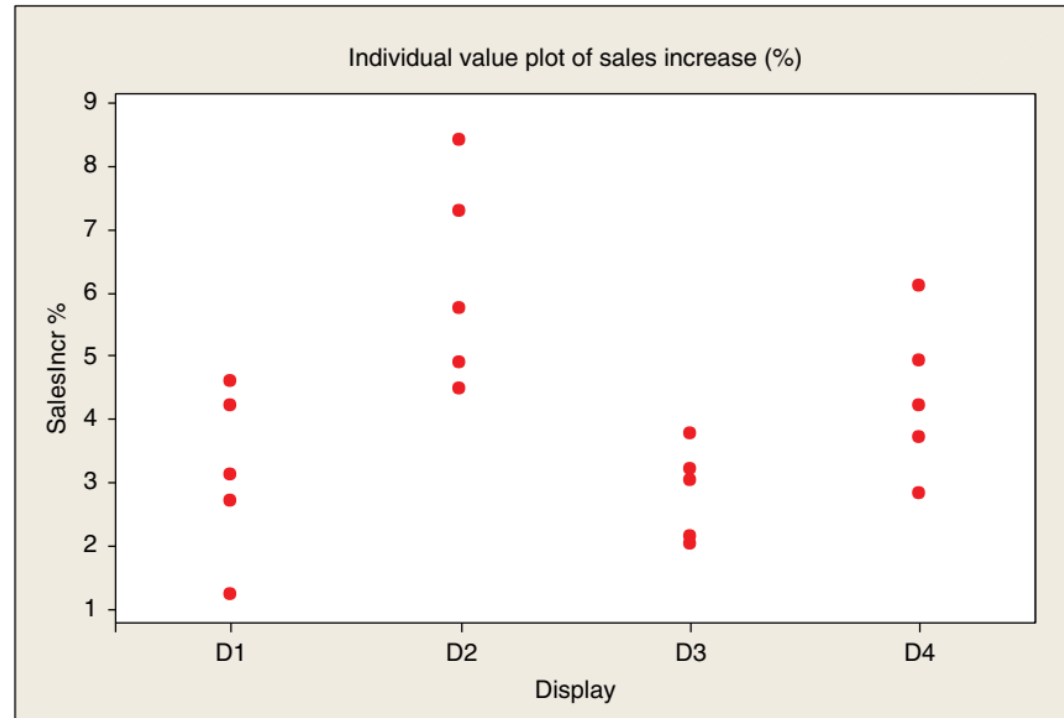
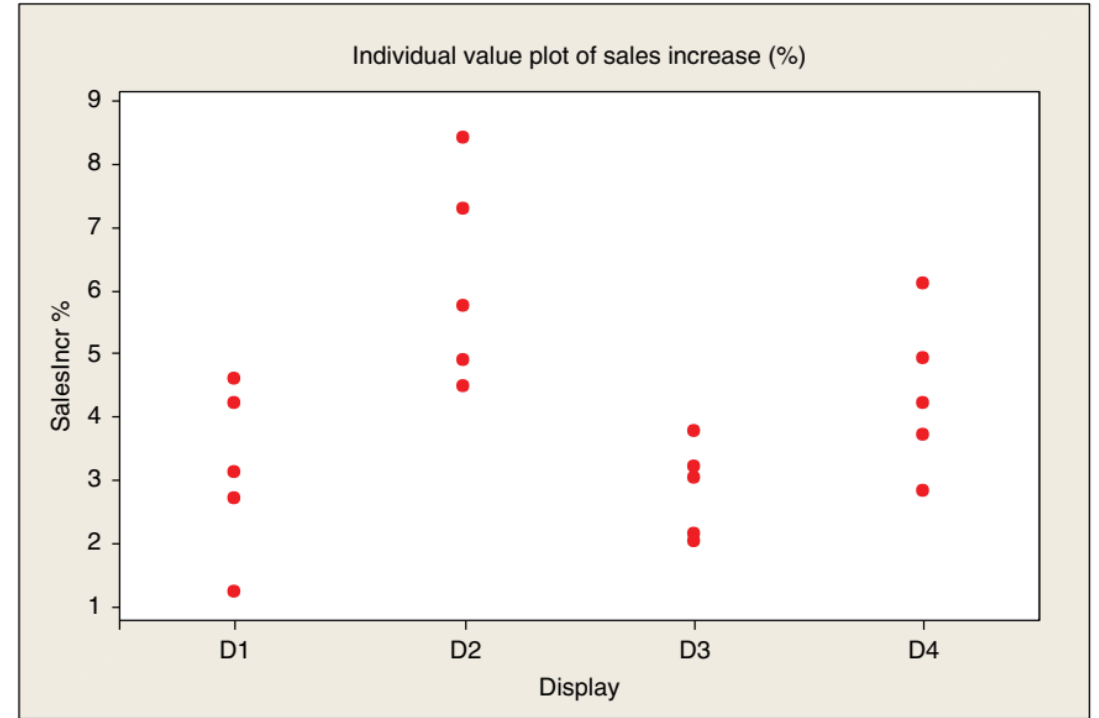


Figure 4.1 Pct. Increase of Shampoo Sales by Display.

- Có sự khác nhau giữa các loại display hay kết quả chỉ là ngẫu nhiên?

Phân tích kết quả

- t-test được dùng cho 2 nhóm
- Tại sao không dùng t-test cho > 2 nhóm
 - Có quá nhiều test phải thực hiện
 - Tỷ lệ lỗi false positive cao
- Phân tích phương sai (analysis of variance - ANOVA) sẽ được dùng để kiểm định > 2 nhóm

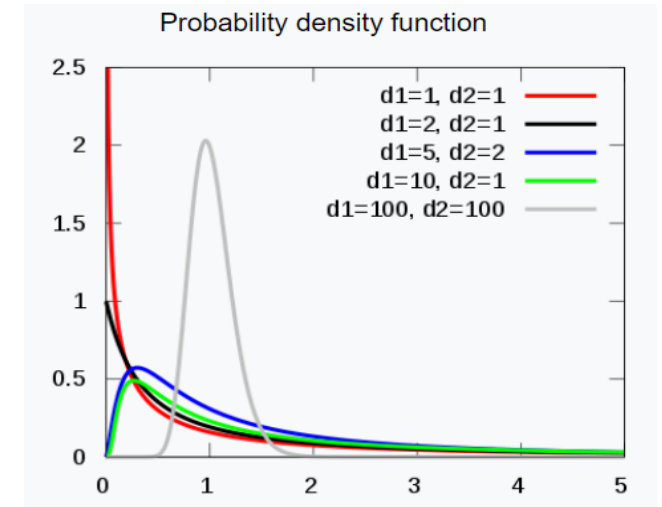
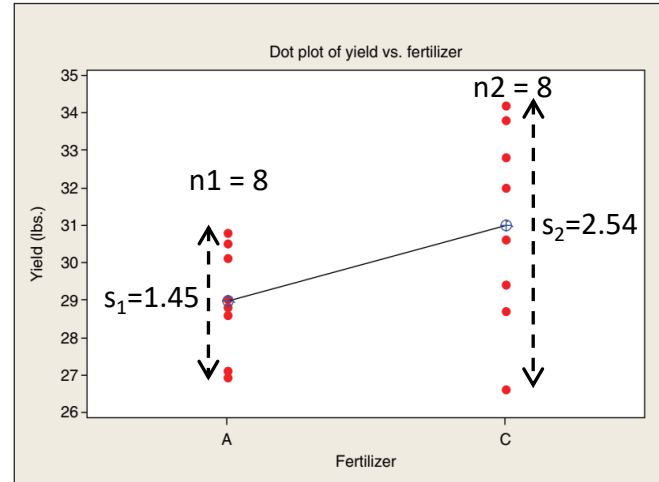


F-test of equality of two variances (nhắc lại slide ngày 09/04)

- $H_0: \sigma_2 = \sigma_1$; $H_1: \sigma_2 \neq \sigma_1$

- Tính $F = \frac{s_2^2}{s_1^2}$

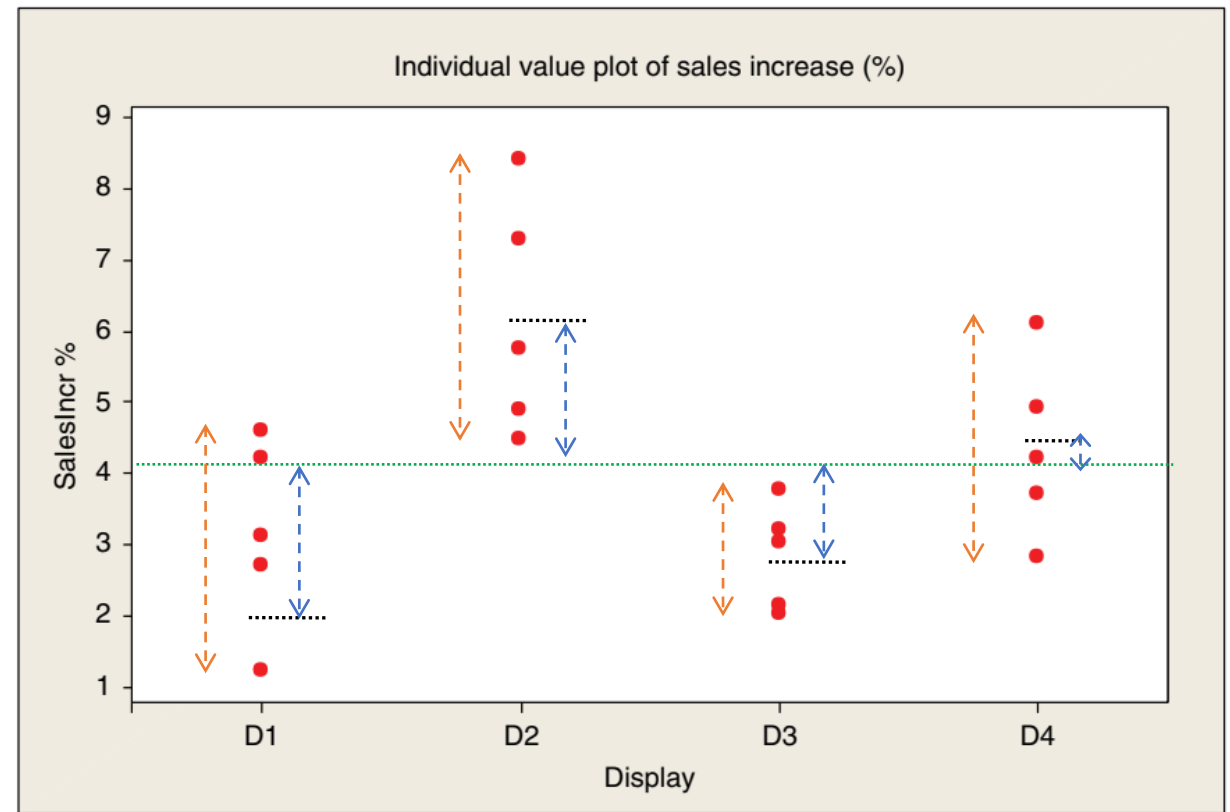
- Tính $P(>F)$



- **Trả lời câu hỏi: có phải 2 mẫu đến từ những quần thể có cùng variance**
- Bác bỏ H_0 : 2 mẫu đến từ những quần thể có cùng variance
- Không bác bỏ H_0 : 2 mẫu đến từ những quần thể khác variance

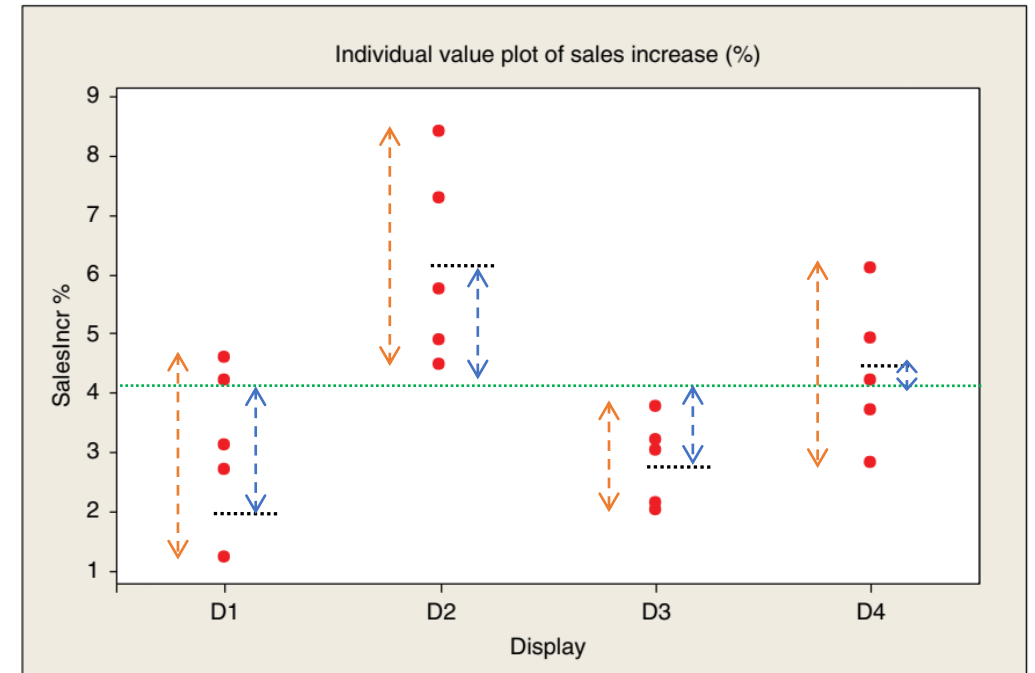
One-way ANOVA F-test

- 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 $P(>F)$



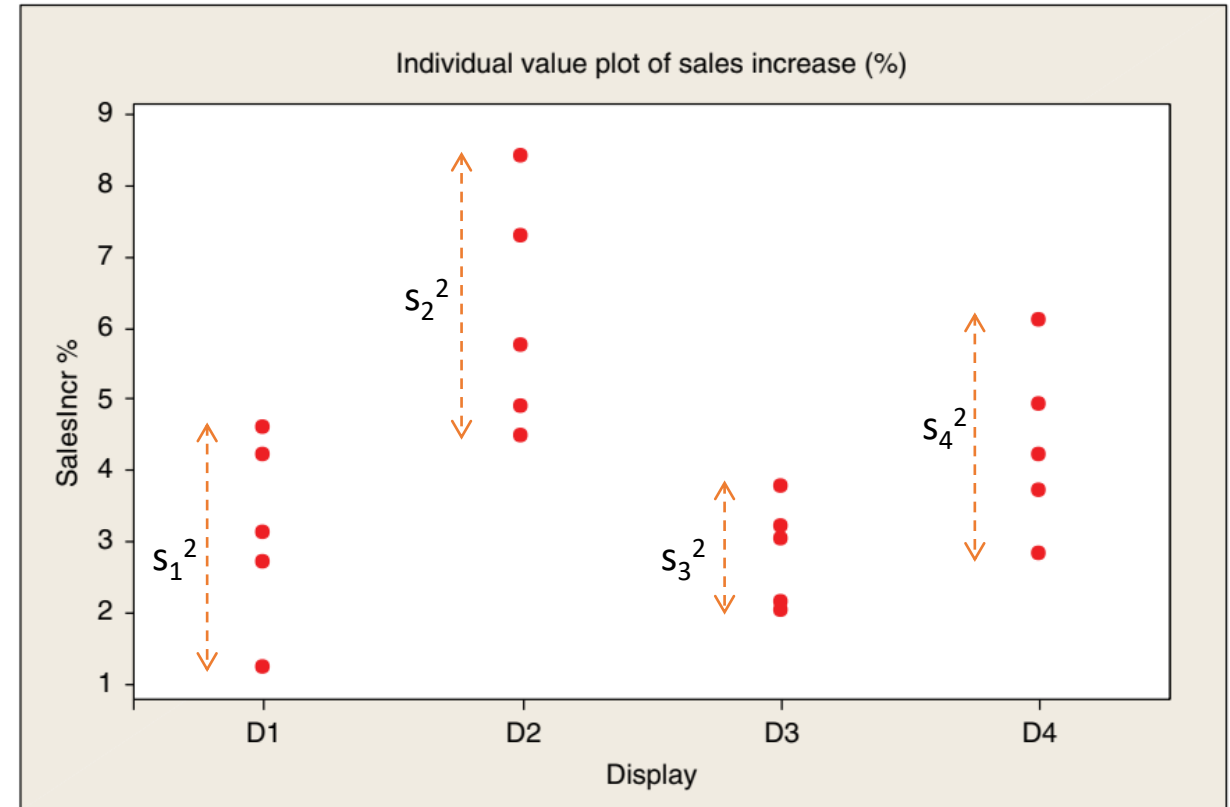
Analysis of variance (ANOVA)

- Có sự khác nhau giữa (mean của) các loại display hay kết quả chỉ là ngẫu nhiên?
 - Nếu không có sự khác biệt giữa **mean** của các nhóm, tất cả sự biến thiên của dữ liệu đều là kết quả của sự biến thiên chung trong toàn bộ quần thể
 - Giả sử variance chung trong toàn bộ quần thể là σ^2
 - Biến thiên của dữ liệu bên trong mỗi nhóm là kết quả của σ^2
 - Biến thiên của dữ liệu giữa các nhóm là kết quả của σ^2
- Within group variance s_w^2 là ước lượng của σ^2
- Among group variance s_a^2 là ước lượng của σ^2

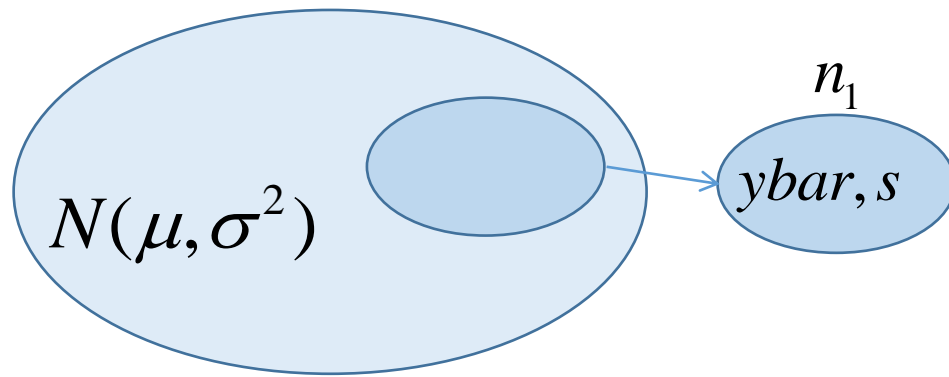


Within group variance s_w^2

- Giả sử các nhóm có cùng phương sai σ^2
- Phương sai gộp (pooled variance) s_w^2 là ước lượng của σ^2 được tính từ phương sai s_i bên trong các nhóm
- Các nhóm cùng kích thước $n_1=n_2=\dots=n_k$, tính average các s_i
 - $s_w^2 = \text{average}(s_1^2, s_2^2, \dots, s_k^2) = \frac{(\sum s_i^2)}{k}$,
 - k là số nhóm, s_i^2 là variance của nhóm i
- Các nhóm khác kích thước, tính weighted average các s_i
 - $s_w^2 = \frac{\sum (n_i - 1)s_i^2}{\sum (n_i - 1)} = \frac{\sum (n_i - 1)s_i^2}{(n - k)}$,
 - n_i là kích thước nhóm i , $n = \sum n_i$ là kích thước tổng



Variance of the mean



- Mean của $ybar$

$$\mu_{ybar} = \mu$$

- Variance của $ybar$

$$\sigma_{ybar}^2 = \frac{\sigma^2}{n_1}$$

- Ước lượng σ^2 từ ước lượng s_{ybar}^2 của σ_{ybar}^2

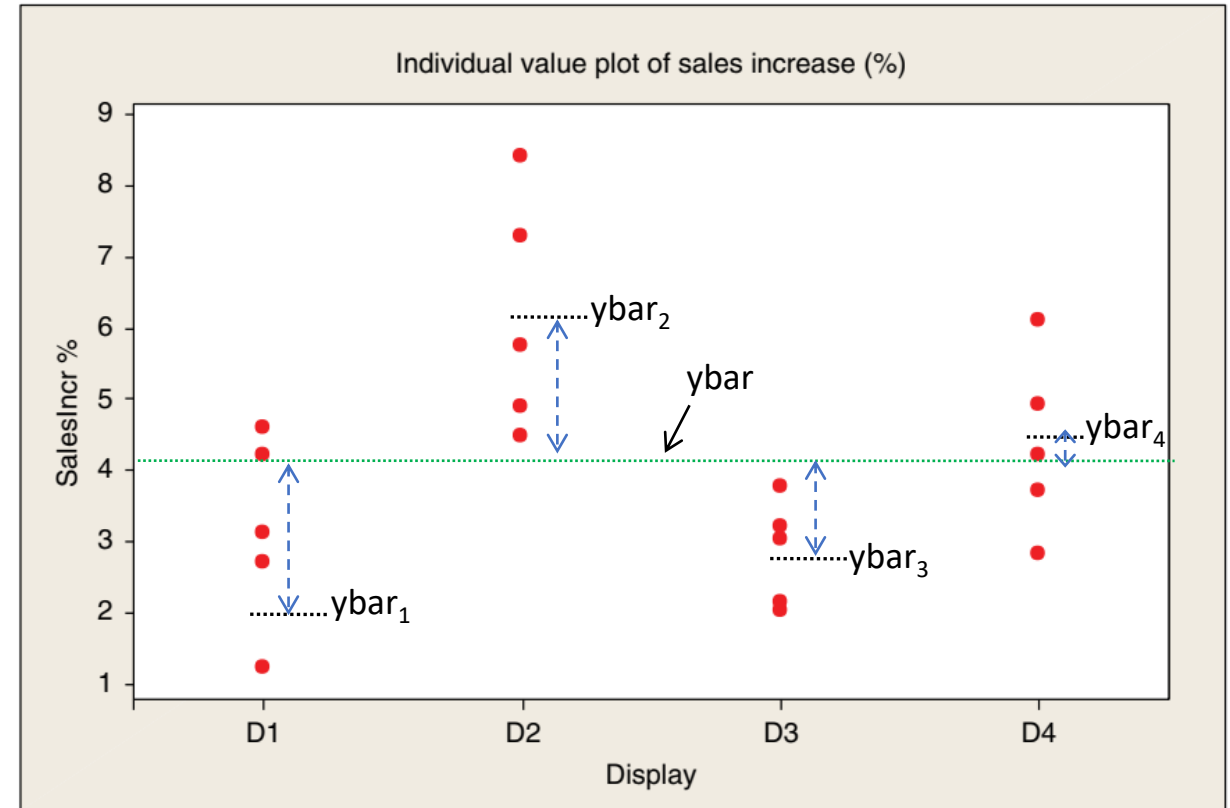
$$\sigma^2 \approx n_1 s_{ybar}^2$$

Among group variance s_a^2

- s_a^2 là ước lượng của phương sai chung σ^2
- Gọi phương sai giữa các giá trị $ybar_i$ của các nhóm là s_{ybar}^2
- Trường hợp các nhóm cùng kích thước $n_1=n_2=...=n_k$
 - $s_{ybar}^2 = \frac{\sum (ybar_i - ybar)^2}{k-1}$,
 - $ybar_i$ là mean của nhóm i, $ybar$ là mean toàn bộ các nhóm
 - $s_a^2 = n_i s_{ybar}^2 = \frac{\sum n_i (ybar_i - ybar)^2}{k-1}$
- Trường hợp các nhóm khác kích thước, $n = \sum n_i$

- $s_{ybar}^2 = \frac{\sum n_i k (ybar_i - ybar)^2}{n(k-1)}$

- $s_a^2 = \frac{n}{k} s_{ybar}^2 = \frac{\sum n_i (ybar_i - ybar)^2}{k-1}$



One-way ANOVA F-test

- H_0 : không có sự khác nhau giữa mean các nhóm
- H_1 : có sự khác nhau giữa mean các nhóm
- s_w^2 và s_a^2 là 2 ước lượng của một phương sai chung σ^2
- Biến ngẫu nhiên $F = \frac{s_a^2}{s_w^2}$ tuân theo phân phối F với
 - $(k - 1)$ bậc tự do trên tử
 - $(n - k)$ bậc tự do dưới mẫu
- Tính p-value tương ứng với F tính được
- Nếu $p\text{-value} < \alpha \rightarrow$ bác bỏ H_0
- Lưu ý: bác bỏ H_0 nghĩa là công nhận có sự khác nhau giữa mean các nhóm chứ **không phải là phương sai các nhóm khác nhau**

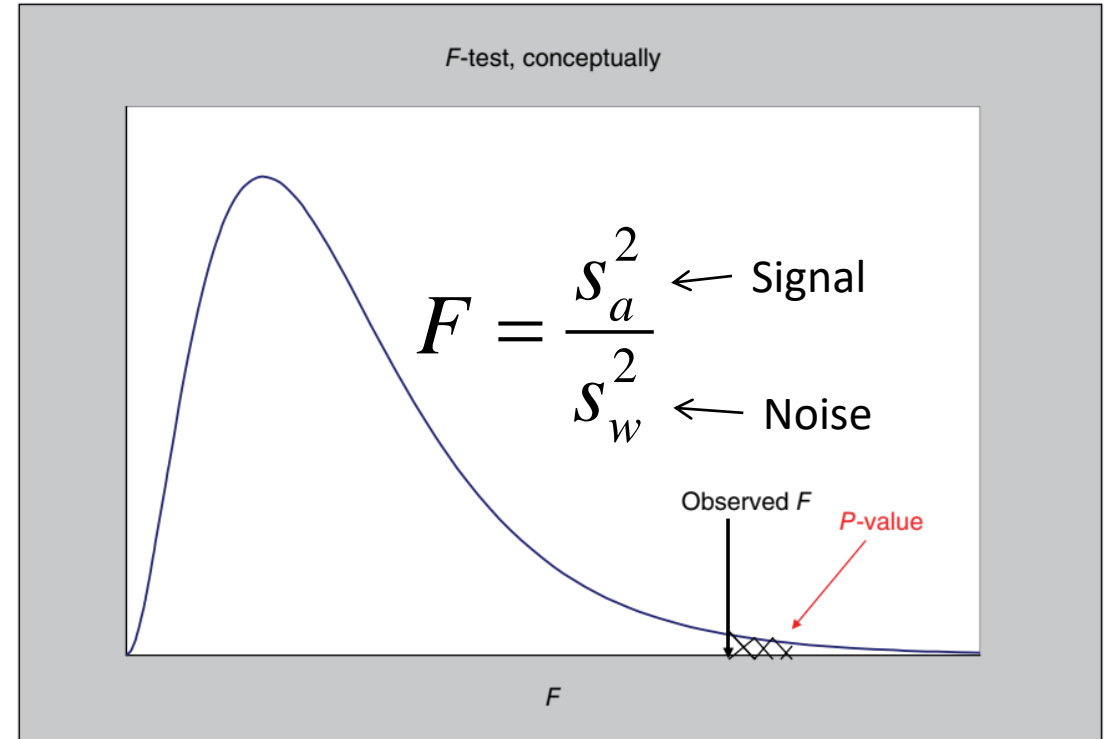


Figure 4.2 Graphical F-Test for Testing the Hypothesis of No Difference Among the Underlying Means for k Treatment Groups.

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	$SS = \sum \sum (y_{ji} - \bar{y})^2$			

ANOVA table

Replication	Displays				
	1	2	3	4	
1	4.2	8.4	3.0	4.9	
2	2.7	4.5	3.8	2.8	
3	3.1	4.9	2.0	6.1	
4	4.6	7.3	2.1	4.2	
5	1.2	5.7	3.2	3.7	
Total reps	5	5	5	5	20
Group mean	3.2	6.2	2.8	4.3	4.1

$$SSA = 5*(3.2-4.1)^2 + 5*(6.2-4.1)^2 + 5*(2.8-4.1)^2 + 5*(4.3-4.1)^2 = 33.91$$

$$SSE = SSW_1 + SSW_2 + SSW_3 + SSW_4 = 26.57$$

$$SSW_1 = (4.2-3.2)^2 + (2.7-3.2)^2 + (3.1-3.2)^2 + (4.6-3.2)^2 + (1.2-3.2)^2 = 7.22$$

$$SSW_2 = (8.4-6.2)^2 + (4.5-6.2)^2 + (4.9-6.2)^2 + (7.3-6.2)^2 + (5.7-6.2)^2 = 10.88$$

$$SSW_3 = (3.0-2.8)^2 + (3.8-2.8)^2 + (2.0-2.8)^2 + (2.1-2.8)^2 + (3.2-2.8)^2 = 2.33$$

$$SSW_4 = (4.9-4.3)^2 + (2.8-4.3)^2 + (6.1-4.3)^2 + (4.2-4.3)^2 + (3.7-4.3)^2 = 6.22$$

ANOVA table

Replication	Displays				
	1	2	3	4	
1	4.2	8.4	3.0	4.9	
2	2.7	4.5	3.8	2.8	
3	3.1	4.9	2.0	6.1	
4	4.6	7.3	2.1	4.2	
5	1.2	5.7	3.2	3.7	
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			

`pf(6.833,3,16,lower.tail = FALSE)=0.00356`

ANOVA trong R

```
> D1 <- c(4.2,2.7,3.1,4.6,1.2)
> D2 <- c(8.4,4.5,4.9,7.3,5.7)
> D3 <- c(3,3.8,2,2.1,3.2)
> D4 <- c(4.9,2.8,6.1,4.2,3.7)
> x=c(D1,D2,D3,D4)
> Displays <- c(rep("D1",5),rep("D2",5),rep("D3",5),rep("D4",5))
> data <- data.frame(x,Displays)
> av <- aov(x ~ Displays)

> summary(av)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Displays	3	34.11	11.369	6.833	0.00356 **
Residuals	16	26.62	1.664		

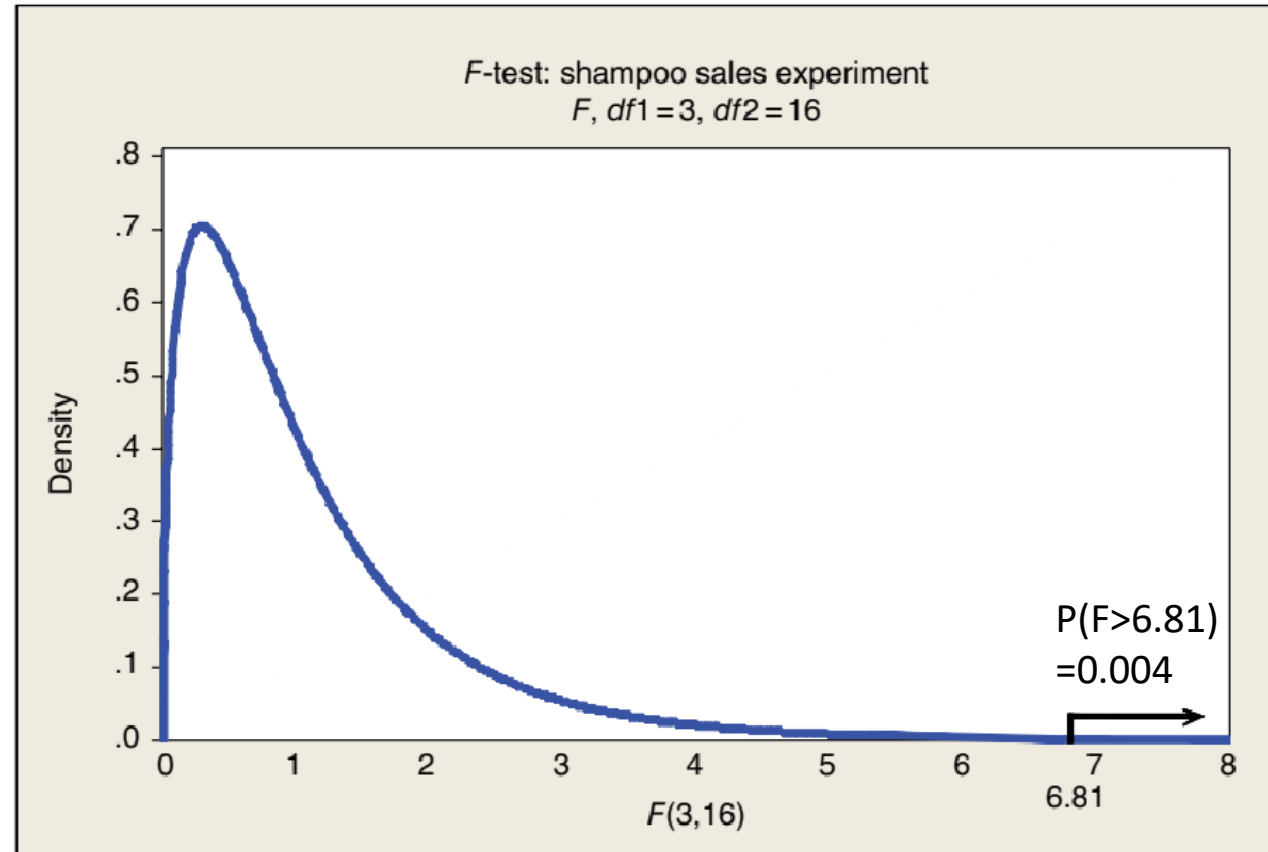


Figure 4.3 Graphical Depiction of F -Test for Shampoo Sales Experiment.

Kết luận: bác bỏ $H_0 \rightarrow$ có sự khác nhau giữa các loại display

Tính khoảng tin cậy 95% 2 bên

Khi chỉ dựa vào dữ liệu 1 nhóm

- Khoảng tin cậy của μ

$$\bar{y}_i \pm t_{.025}(n_i - 1)s_p / \sqrt{n_i}$$

- Khoảng tin cậy của trung bình $\mu_1 - \mu_2$

$$\bar{y}_2 - \bar{y}_1 \pm t_{.025}(n_1 + n_2 - 2)s_p / \sqrt{1/n_1 + 1/n_2}$$

Khi chỉ dựa vào dữ liệu nhiều nhóm

- Khoảng tin cậy của μ

$$\bar{y}_i \pm t_{.025, n-k} \times S / \sqrt{n_i} \quad \text{với} \quad S = \sqrt{s_w^2} = \sqrt{MSE}$$

- Khoảng tin cậy của trung bình $\mu_1 - \mu_2$

$$\bar{y}_1 - \bar{y}_2 \pm t_{.025}(n - k)S / \sqrt{1/n_1 + 1/n_2}$$

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

T-test vs F-test

Thử nghiệm 2 loại phân bón A và B trên 20 cây cà chua được kết quả sau

A	30.5	28.8	28	27.1	30.1	30.8	26.9	28.8	30	29
B	32	26.6	34.2	28.7	32.8	30.6	32.8	29.4	31.9	31



$$H_0: \mu_B - \mu_A = 0$$

$$H_0: \mu_B - \mu_A > 0$$

t	df	p-value
2.3962	18	.0138

$$s_p = 1.866$$

$$H_0: \mu_B - \mu_A = 0$$

$$H_0: \mu_B - \mu_A \neq 0$$

t	df	p-value
2.3962	18	.0276

$$s_p = 1.866$$

$$H_0: \mu_B - \mu_A = 0 \quad ; \quad H_0: \mu_B - \mu_A \neq 0$$

Source	df	SS	MS	F	P-value
Fertilizers	1	20.0	20.0	5.742	.0276
Error	18	62.7	3.483		
Total	19	82.7			

$$S = \sqrt{3.483} = 1.866$$