XÁC SUẤT - THỐNG KÊ

CHƯƠNG 6: PHÂN TÍCH PHƯƠNG SAI

TS. Phan Thị Hường

Trường Đại học Bách Khoa TP HCM
Khoa Khoa học ứng dụng, bộ môn Toán ứng dụng
Email: huongphan@hcmut.edu.vn

TP. HCM — 2020.

Nội dung

1. Phân tích phương sai một nhân tố

NỘI DUNG

1. PHÂN TÍCH PHƯƠNG SAI MỘT NHÂN TỐ

2. PHÂN TÍCH PHƯƠNG SAI HAI NHÂN TỐ

GIỚI THIỆU

VÍ DỤ 1.1

Một nhà máy sản xuất bao bì quan tâm đến việc tăng độ đàn hồi của các túi giấy do nhà máy làm ra. Các kỹ sư của nhà máy cho trằng độ đàn hồi của các túi giấy bị ảnh hưởng bởi hàm lượng gỗ cứng trong nguyên liệu, phạm vi thay đổi được quan tâm là từ 5% *đến* 20%. *Các kỹ sư quyết định thử nghiệm với hàm lượng gỗ cứng trong bột gỗ ở* 4 *mức:* 5%, 10%, 15% *và* 20%. *Ở mỗi mức,* 6 *mẫu vật được chọn để kiểm trong trong phòng thí nghiệm, theo thứ tự ngẫu nhiên. Kết quả cho bởi bảng bên dưới:*

GIỚI THIỆU

Hàm lượng gỗ cứng (%)	Quan trắc 1	2	3	4	5	6	Tổng	Trung bình
5	7	8	15	11	9	10	60	10.00
10	12	17	13	18	19	15	94	15.67
15	14	18	19	17	16	18	102	17.00
20	19	25	22	23	18	20	127	21.17
							383	15.96

BẢNG: Độ đàn hồi của các bao bì giấy

GIỚI THIỆU

- Câu hỏi đặt ra là: có sự khác biệt về độ đàn hồi (psi) giữa các sản phẩm có hàm lượng gỗ cứng trong bột gỗ ở 4 mức khác nhau hay không?
- Ví dụ trên đặt ra bài toán so sánh sự khác biệt giữa trung bình của nhóm khác nhau (≥ 3).
- Thí nghiệm trên được gọi là Thí nghiệm ngẫu nhiên đầy đủ với một nhân tố (The completely Randomized Single-Factor).
- Để trả lời câu hỏi, ta sử dụng kỹ thuật Phân tích phương sai (Analysis of Variance - ANOVA).

BÀI TOÁN PHÂN TÍCH PHƯƠNG SAI

- Giả sử ta cần so sánh k mức khác nhau của một nhân tố. Mỗi mức của nhân tố được gọi là một phương thức xử lý (treatment). Kết quả của mỗi một phương thức làm một biến ngẫu nhiên. Dữ liệu quan trắc được sẽ được biểu diễn giống như trong bảng 1 và bảng 2, mỗi giá trị trong bảng 1, ký hiệu là y_{ij}, gọi là quan trắc thứ j được chọn dưới phương thức xử lý i. Giả sử ở mỗi phương thức xử lý, ta chọn số quan trắc bằng nhau, bằng n.

Bài toán Phân tích phương sai

Bảng dữ liệu tổng quát cho mô hình (2) là

Treatment	Observations				Totals	Averages
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	y_{k1}	y_{k2}	$\cdots$	y_{kn}	$y_{k\cdot}$	$\bar{y}_{k\cdot}$
					$y_{\cdot\cdot}$	$\bar{y}_{\cdot\cdot}$

Trong đó,

$$y_{i\cdot} = \sum_{j=1}^{n} y_{ij}, \quad \bar{y}_{i\cdot} = y_{i\cdot}/n, \qquad i = 1, 2, \ldots, k$$

$$y_{\cdot\cdot} = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}, \quad \bar{y}_{\cdot\cdot} = y_{\cdot\cdot}/N, \quad N = kn$$

BÀI TOÁN PHÂN TÍCH PHƯƠNG SAI

- Dữ liệu trong bảng 1 có thể được biểu diễn theo mô hình tuyến tính

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \tag{1}$$

với $i = 1, 2, \ldots, k$ và $j = 1, 2, \ldots, n$. Trong đó:

- μ là giá trị trung bình chung,
- τ_i là ảnh hưởng của phương thức xử lý thứ i và ϵ_{ij} là thành phần sai số.

Bài toán Phân tích phương sai

- Mô hình (1) được viết lại như sau

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{2}$$

với μ_i là trung bình của phương thức xử lý thứ i.

- Các giả định của mô hình:

Bài toán Phân tích phương sai

- Mô hình (1) được viết lại như sau

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{2}$$

  với μ_i là trung bình của phương thức xử lý thứ i.
- Các giả định của mô hình:
  - Tổng thể có phân phối chuẩn,

BÀI TOÁN PHÂN TÍCH PHƯƠNG SAI

- Mô hình (1) được viết lại như sau

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{2}$$

với μ_i là trung bình của phương thức xử lý thứ i.

- Các giả định của mô hình:
  - Tổng thể có phân phối chuẩn,
  - Tổng thể có phương sai bằng nhau, suy ra $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$,

Bài toán Phân tích phương sai

- Mô hình (1) được viết lại như sau

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{2}$$

với μ_i là trung bình của phương thức xử lý thứ i.

- Các giả định của mô hình:
    - Tổng thể có phân phối chuẩn,
    - Tổng thể có phương sai bằng nhau, suy ra $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$,
    - Mẫu phải được chọn ngẫu nhiên và độc lập.

Bài toán Phân tích phương sai

- Mô hình (1) được viết lại như sau

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{2}$$

với μ_i là trung bình của phương thức xử lý thứ i.

- Các giả định của mô hình:
  - Tổng thể có phân phối chuẩn,
  - Tổng thể có phương sai bằng nhau, suy ra $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$,
  - Mẫu phải được chọn ngẫu nhiên và độc lập.
  - Đối với mô hình với những hiệu ứng cố định (fixed-effects model), điều kiện cho các $\tau_i, i = 1, 2, \ldots, k$ là

$$\sum_{i=1}^{k} \tau_i = 0 \tag{3}$$

GIẢ THUYẾT CỦA BÀI TOÁN PHÂN TÍCH PHƯƠNG SAI MỘT NHÂN TỐ

- Giả thuyết:

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0$$

Trung bình của tất cả các phương thức xử lý bằng nhau, hay nói cách khác, không có sự khác biệt về trung bình giữa các nhóm.

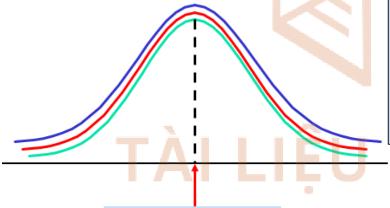
- Đối thuyết:

$$H_1 : \tau_i \neq 0 \quad \text{với ít nhất một } i$$

Nghĩa là có tồn tại sự khác biệt giữa các nhóm, nhưng không có nghĩa là tất cả trung bình đều khác nhau (có thể có một vài cặp).

GIẢ THUYẾT VÀ ĐỐI THUYẾT

$$\begin{cases} H_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0 \\ H_1 : \tau_i \neq 0 \quad \text{với ít nhất một } i \end{cases}$$



Tất cả trung bình bằng nhau:

Giả thuyết H_0 đúng

(Không có sự khác biệt giữa các nhóm)

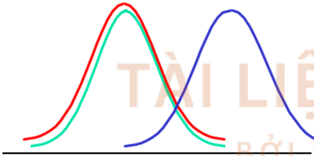
$\mu_1 = \mu_2 = \mu_3$

GIẢ THUYẾT VÀ ĐỐI THUYẾT

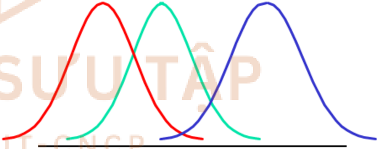
$$\begin{cases} H_0 : H_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0 \\ H_1 : \tau_i \neq 0 \quad \text{với ít nhất một } i \end{cases}$$

Giả thuyết H_0 sai

(Có sự khác biệt giữa các nhóm)



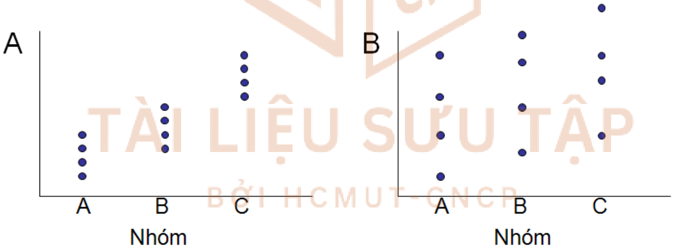
hoặc



$\mu_1 = \mu_2 \neq \mu_3$ $\qquad$ $\mu_1 \neq \mu_2 \neq \mu_3$

Sự biến thiên

- Sự biến thiên trong dữ liệu là chìa khóa để kiểm tra sự bằng nhau của trung bình giữa các nhóm.
- Ví dụ: trong mỗi trường hợp bên dưới, các giá trị trung bình nhìn có vẻ khác nhau, nhưng sự biến thiên lớn trong các nhóm ở B là bằng chứng cho thấy rằng sự khác nhau giữa các trung bình là rất nhỏ.



Biến thiên nhỏ trong các nhóm

Biến thiên lớn trong các nhóm

PHÂN CHIA SỰ BIẾN THIÊN

Sự biến thiên toàn phần trong dữ liệu có thể phân chia thành hai thành phần như sau

$$SST = SSW + SSB \tag{4}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 + n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2$$

Trong đó

- SST = Tổng bình phương toàn phần (Total Sum of Squares).
- SSW = Tổng bình phương bên trong các các nhóm (Sum of Squares Within Groups).
- SSB = Tổng bình phương giữa các nhóm (Sum of Squares Between Groups)

CÔNG THỨC RÚT GỌN

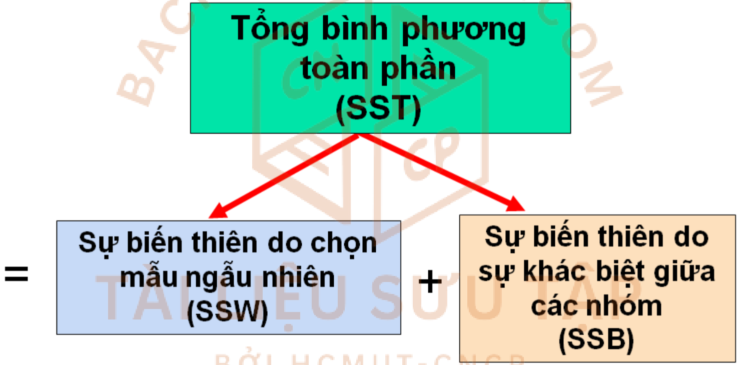
Tổng các bình phương của ANOVA với cỡ mẫu bằng nhau trong mỗi phương thức xử lý thường được tính bởi các công thức rút gọn sau

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{5}$$

$$SSB = \sum_{i=1}^{k} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N} \tag{6}$$

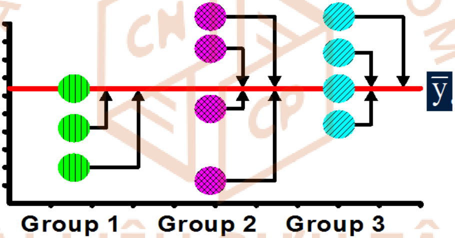
$$SSW = SST - SSB \tag{7}$$

PHÂN CHIA SỰ BIẾN THIÊN



SỰ BIẾN THIÊN TOÀN PHẦN

$$SST = (y_{11} - \bar{y}_{\bullet\bullet})^2 + (y_{12} - \bar{y}_{\bullet\bullet})^2 + \ldots + (y_{kn} - \bar{y}_{\bullet\bullet})^2$$

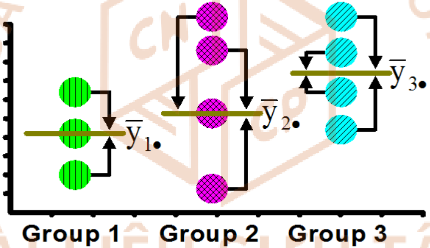


Trung bình bình phương toàn phần :

$$MST = \frac{SST}{kn - 1}$$

SỰ BIẾN THIÊN TRONG TỪNG NHÓM

$$SSW = (y_{11} - \bar{y}_{1\bullet})^2 + (y_{12} - \bar{y}_{1\bullet})^2 + \ldots + (y_{kn} - \bar{y}_{k\bullet})^2$$

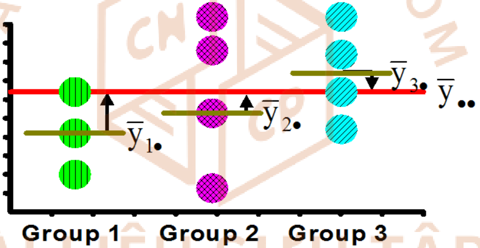


Trung bình bình phương trong từng nhóm

$$MSW = \frac{SSW}{k(n-1)}$$

SỰ BIẾN THIÊN GIỮA CÁC NHÓM

$$SSB = n(\bar{y}_{1\bullet} - \bar{y}_{\bullet\bullet})^2 + n(\bar{y}_{2\bullet} - \bar{y}_{\bullet\bullet})^2 + \ldots + n(\bar{y}_{k\bullet} - \bar{y}_{\bullet\bullet})^2$$



Trung bình bình phương giữa các nhóm

$$MSB = \frac{SSB}{k-1} \tag{8}$$

Trung bình bình phương

- Trung bình bình phương toàn phần

$$MST = \frac{SST}{kn-1} \tag{9}$$

- Trung bình bình phương trong từng nhóm

$$MSW = \frac{SSW}{k(n-1)} \tag{10}$$

- Trung bình bình phương giữa các nhóm

$$MSB = \frac{SSB}{k-1} \tag{11}$$

ANOVA một nhân tố và Thống kê F

Mệnh đề 1.1

Kỳ vong của tổng bình phương giữa các nhóm:

$$E(SSB) = (k-1)\sigma^2 + n\sum_{i=1}^{k}\tau_i^2$$

Kỳ vọng của tổng bình phương trong các nhóm:

$$E(SSW) = a(n-1)\sigma^2$$

ANOVA một nhân tố và Thống kê F

$$\begin{cases} H_0 : H_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0 \\ H_1 : \tau_i \neq 0 \quad \text{với ít nhất một } i \end{cases}$$

- Thống kê kiểm định

$$F = \frac{MSB}{MSW} \tag{12}$$

Thống kê F là tỷ số giữa phương sai ước lượng giữa các nhóm với phương sai ước lượng trong từng nhóm.

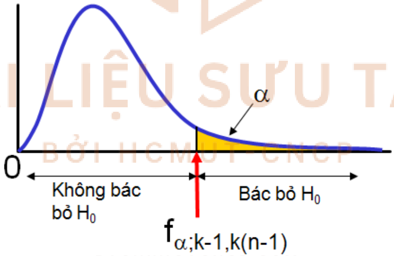
- Bậc tự do
  - $df_1 = k - 1$ (k = số nhóm)
  - $df_2 = k(n-1)$ ($kn = N$: tổng số phần tử khảo sát ở tất cả các nhóm)

BẢNG ANOVA MỘT NHÂN TỐ

Nguồn của sự biến thiên	SS	df	MS	F
Giữa các nhóm	SSB	$k-1$	MSB	$F = \frac{MSB}{MSW}$
Trong từng nhóm	SSW	$k(n-1)$	MSW	
Tổng	SST	$kn-1$		

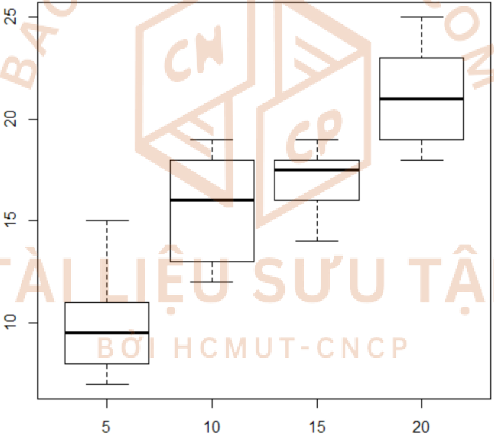
- Bác bỏ H_0 khi:

$$f > f_{\alpha;k-1,k(n-1)} \tag{13}$$



VÍ DỤ

VÍ DỤ 1. (tiếp theo)
Đồ thị boxplot cho so sánh trung bình:



Ví dụ

Với $\alpha = 0.01$, ta kiểm định giả thuyết

$$\begin{cases} H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0 \\ H_1 : \tau_i \neq 0 \text{ với ít nhất một } i \end{cases}$$

Tính các tổng bình phương liên quan, từ bảng 1 ta có

$$SST = \sum_{i=1}^{4}\sum_{j=1}^{6} y_{ij}^2 - \frac{y_{..}^2}{N} = \left(7^2 + 8^2 + \ldots + 20^2\right) - \frac{383^2}{24} = 512.96$$

$$SSB = \sum_{i=1}^{4} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N} = \frac{60^2 + 94^2 + 102^2 + 127^2}{6} - \frac{383^2}{24} = 382.79$$

$$SSW = SST - SSB = 512.96 - 382.79 = 130.17$$

Ví dụ

- Tính giá trị thống kê kiểm định:

$$f = \frac{MSB}{MSW} = \frac{SSB/(k-1)}{SSW/(kn-k)} = \frac{382.79/3}{130.17/20} = 19.60$$

Nguồn của sự biến thiên	SS	df	MS	F
Giữa các nhóm (SSB)	382.79	3	127.5967	
Trong từng nhóm (SSW)	130.17	20	6.5085	$F = 19.60$
Tổng	130.17	23		

Vì $f_{0.01;3,20} = 4.94$, nên ta bác bỏ H_0 và kết luận rằng hàm lượng gỗ cứng trong nguyên liệu có ảnh hưởng đến độ đàn hồi của các bao bì giấy.
Kết luận dùng p-giá trị:

$$p = \mathbb{P}(F_{3,20} > 19.60) = 3.59 \times 10^{-6}$$

vì $p = 3.59 \times 10^{-6} << \alpha$, nên ta có đủ bằng chứng mạnh để bác bỏ H_0.

BÀI TẬP 1.1

Năm 1992, trong một nghiên cứu về ảnh hưởng của trục cuốn ép lên cường độ chịu nén của các loại thùng cảton tiêu chuẩn RSC được sản xuất, Burgess đã tiến hành đo cường độ chịu nén của bốn loại thùng carton khác nhau. Dưới đây là dữ liệu thu được

Hộp	Lực nén					
1	655.5	688.3	734.3	721.4	679.1	699.4
2	789.2	772.5	786.9	686.1	732.1	774.8
3	737.1	639.0	727.1	671.7	717.2	727.1
4	535.1	628.7	542.4	559.0	586.9	520.0

Hãy so sánh cường độ lực nén của bốn loại thùng carton với mức ý nghĩa $\alpha = 0.01$.

ANOVA với cỡ mẫu không bằng nhau

- Khi cỡ mẫu giữa các phương thức xử lý không bằng nhau, các công thức tính tổng bình phương cần phải hiệu chỉnh lại. Xét bài toán ANOVA với k phương thức xử lý, với phương thức thứ $i (i = 1, 2, \ldots, k)$, chọn mẫu n_i phần tử. Tổng số phần tử là $N = \sum_{i=1}^{k} n_i$. Công thức tính SST, SSB và SSW được hiệu chỉnh lại như sau:

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{14}$$

$$SSB = \sum_{i=1}^{k} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} \tag{15}$$

$$SSW = SST - SSB \tag{16}$$

So sánh bội sau ANOVA

- Khi giả thuyết $H_0 : \tau_1 = \tau_2 = \ldots = \tau_k$ được bác bỏ trong ANOVA, ta biết có sự khác biệt giữa các nhóm (các phương thức thí nghiệm khác nhau); nhưng ANOVA không chỉ rõ nhóm nào gây ra sự khác biệt.
- Để xác định trung bình của nhóm nào là khác biệt, ta sử dụng phương pháp so sánh bội (Multiple comparison method). Một phương pháp so sánh bội đơn giản là Phương pháp ý nghĩa độ lệch nhỏ nhất (Least significant difference - LSD) của Fisher.
- Nội dung của phương pháp LSD là so sánh tất cả các cặp giá trị trung bình với giả thuyết $H_0 : \mu_i = \mu_j$ (với mọi $i \neq j$), sử dụng thống kê t

$$t = \frac{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}}{\sqrt{\frac{2MSW}{n}}} \tag{17}$$

SO SÁNH BỘI SAU ANOVA

- Giả sử với đối thuyết hai phía, cặp giá trị trung bình μ_i và μ_j gọi là khác nhau có ý nghĩa nếu

$$|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}| > LSD$$

với

$$LSD = t_{1-\alpha/2}^{k(n-1)} \sqrt{\frac{2MSW}{n}} \tag{18}$$

- Nếu mỗi phương thức thí nghiệm có cỡ mẫu khác nhau, LSD được định nghĩa như sau

$$LSD = t_{1-\alpha/2}^{N-k} \sqrt{MSW\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

trong đó: N là tổng số phần tử khảo sát, k là số nhóm, n_i và n_j lần lượt là số phần tử của phương thức thí nghiệm thứ i và j.

Ví dụ

Thực hiện so sánh bội với kết quả phân tích ANOVA trong ví dụ 1, ta có số nhóm $k = 4$, $n = 6$, $MSE = 6.51$ và $t^{20}_{0.975} = 2.086$; trung bình ở các phương thức xử lý như sau

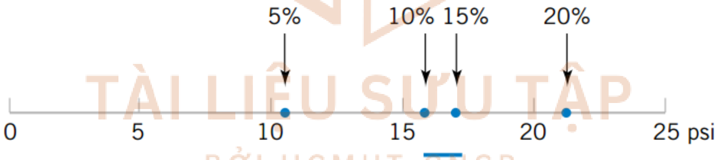
$$\bar{y}_{1.} = 10.00; \quad \bar{y}_{2.} = 15.67; \quad \bar{y}_{3.} = 17.00; \quad \bar{y}_{4.} = 21.17$$

Giá trị $LSD = t^{20}_{0.975}\sqrt{2MSE/n} = 2.086\sqrt{2 \times 6.51/6} = 3.07$. So sánh các cặp giá trị trung bình như sau:

$$4\ vs.\ 1 = 21.17 - 10.00 = 11.17 > 3.07$$
$$4\ vs.\ 2 = 21.17 - 15.67 = 5.50 > 3.07$$
$$4\ vs.\ 3 = 21.17 - 17.00 = 4.17 > 3.07$$
$$3\ vs.\ 1 = 17.00 - 10.00 = 7.00 > 3.07$$
$$3\ vs.\ 2 = 17.00 - 15.67 = 1.33 < 3.07$$
$$2\ vs.\ 1 = 15.67 - 10.00 = 5.67 > 3.07$$

VÍ DỤ

Từ phân tích trên, ta nhận thấy rằng ngoại trừ cặp giá trị trung bình của nhóm 2 và 3, tất cả các cặp còn lại đều khác nhau có ý nghĩa. Điều này chứng tỏ rằng hàm lượng gỗ cứng trong bột gỗ ở mức 10% và 15% đều cho ra những sản phẩm có độ đàn hồi xấp xỉ bằng nhau. Kết quả có thể được mô tả bởi hình vẽ sau:



GIỚI THIỆU ANOVA 2 NHÂN TỐ

ANOVA hai nhân tố dùng để:

- Nghiên cứu tác động của
  - Hai nhân tố được quan tâm trên một biến phụ thuộc (biến giải thích).
    Ví dụ: ảnh hưởng của ánh sáng (cường độ: mạnh, trung bình, yếu) và lượng nước (nhiều, ít) tưới lên chiều cao của cây.
  - Tương tác giữa các mức khác nhau của hai nhân tố.
    Chẳng hạn, có tương tác nào giữa cường độ ánh sáng mạnh chiếu lên cây khi tưới nhiều nước hay không?

CÁC GIẢ ĐỊNH CỦA MÔ HÌNH

Bài toán ANOVA hai nhân tố được biểu diễn theo mô hình thống kê tuyến tính như sau:

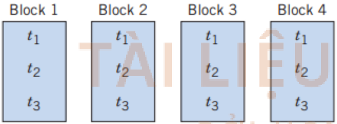
$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad \begin{cases} i & = 1, 2, \dots, a \\ j & = 1, 2, \dots, b \end{cases} \tag{19}$$

với μ là trung bình chung, τ_i là ảnh hưởng của phương thức xử lý thứ i (của nhân tố A), β_j là ảnh hưởng của khối thứ j (của nhân tố B); ϵ_{ij} là thành phần sai số.

- Các giả định:
  - Tổng thể có phân phối chuẩn,
  - Tổng thể có phương sai bằng nhau,
  - Mẫu ngẫu nhiên được chọn độc lập,

THIẾT KẾ VÀ PHÂN TÍCH

- Đối với bài toán ANOVA 2 nhân tố, ta sử dụng phương pháp thiết kế gọi là Thiết kế khối ngẫu nhiên đầy đủ (Randomized complete block design). Phương pháp là chọn b khối (block) ứng với b mức của nhân tố B và lặp lại thí nghiệm đầy đủ với a mức tương ứng của nhân tố A. Như vậy sẽ có a quan trắc ứng với mỗi khối, với thứ tự chọn một cách ngẫu nhiên bên trong khối đó. Mô tả phương pháp như hình vẽ bên dưới:



A randomized complete block design.

A Randomized Complete Block Design

Treatments (Method)	Block (Girder) 1	2	3	4
1	y_{11}	y_{12}	y_{13}	y_{14}
2	y_{21}	y_{22}	y_{23}	y_{24}
3	y_{31}	y_{32}	y_{33}	y_{34}

THIẾT KẾ VÀ PHÂN TÍCH

- Bảng dữ liệu cho ANOVA 2 nhân tố:

A Randomized Complete Block Design with a Treatments and b Blocks

	Blocks					
Treatments	1	2	...	b	Totals	Averages
1	y_{11}	y_{12}	...	y_{1b}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	...	y_{2b}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	...	y_{ab}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
Totals	$y_{\cdot 1}$	$y_{\cdot 2}$	...	$y_{\cdot b}$	$y_{\cdot\cdot}$	
Averages	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	...	$\bar{y}_{\cdot b}$		$\bar{y}_{\cdot\cdot}$

PHÂN CHIA SỰ BIẾN THIÊN

- SST = SSG + SSB + SSE

Tổng bình phương toàn phần (SST) =

Sự biến thiên gây ra bởi sự khác nhau giữa các nhóm (SSG)

\+

Sự biến thiên gây ra bởi sự khác nhau giữa các khối (SSB)

\+

Sự biến thiên do chọn mẫu ngẫu nhiên (Sai số không giải thích được) (SSE)

Sai số được giả sử độc lập, có phân phối chuẩn và phương sai bằng nhau →

TỔNG CÁC BÌNH PHƯƠNG

- Định nghĩa tổng các bình phương như sau:

$$\text{Toàn phần}: \; SST = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2 \tag{20}$$

$$\text{Giữa các nhóm}: \; SSG = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 \tag{21}$$

$$\text{Giữa khác khối}: \; SSB = a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2 \tag{22}$$

$$\text{Sai số}: \; SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y}_{..})^2 \tag{23}$$

- Bậc tự do tương ứng:

$$ab - 1 = (a-1) + (b-1) + (a-1)(b-1)$$

Trung bình bình phương

Các trung bình bình phương được định nghĩa như sau:

$$MST = \frac{SST}{ab-1} \tag{24}$$

$$MSG = \frac{SSG}{a-1} \tag{25}$$

$$MSB = \frac{SSB}{b-1} \tag{26}$$

$$MSE = \frac{SSE}{(a-1)(b-1)} \tag{27}$$

GIẢ THUYẾT CỦA BÀI TOÁN ANOVA 2 NHÂN TỐ

- Đối với các nhóm (nhân tố A):

$$\begin{cases} H_{0a}: \tau_1 = \tau_2 = \ldots = \tau_a = 0 \\ H_{1a}: \tau_i \neq 0 \text{ với ít nhất một } i \end{cases} \tag{28}$$

  - Thống kê kiểm định F:

$$F_a = \frac{MSG}{MSE} \tag{29}$$

  - Bác bỏ H_{0a} khi:

$$f_{0a} > f_{\alpha;a-1,(a-1)(b-1)}$$

GIẢ THUYẾT CỦA BÀI TOÁN ANOVA 2 NHÂN TỐ

- Đối với các khối (nhân tố B):

$$\begin{cases} H_{0b}: \beta_1 = \beta_2 = \ldots = \beta_b = 0 \\ H_{1b}: \beta_j \neq 0 \text{ với ít nhất một } j \end{cases} \tag{30}$$

  - Thống kê kiểm định F:

$$F_b = \frac{MSB}{MSE} \tag{31}$$

  - Bác bỏ H_{0b} khi:

$$f_{0b} > f_{\alpha;b-1,(a-1)(b-1)}$$

BẢNG ANOVA 2 NHÂN TỐ

Nguồn của sự biến thiên	SS	df	MS	F
Giữa các nhóm	SSG	$a-1$	MSG	$F_a = \frac{MSG}{MSE}$
Giữa các khối	SSB	$b-1$	MSB	$F_b = \frac{MSB}{MSE}$
Sai số	SSE	$(a-1)(b-1)$	MSE	
Tổng	SST	$ab-1$		

Nhiều hơn một quan trắc trong một ô

- Một thiết kế ANOVA 2 chiều có nhiều hơn một quan trắc trong mỗi ô làm tăng nguồn gây ra sự biến thiên.
- Xuất hiện tương tác giữa các nhóm và các khối.
- Gọi
  - a = số nhóm (nhân tố A),
  - b = số khối (nhân tố B),
  - n = số quan trắc trong mỗi ô,
  - $N = abn$ = tổng số quan trắc ứng với ab ô.
- Mô hình được biểu diễn như sau

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \quad \begin{cases} i & = 1,2,\ldots,a \\ j & = 1,2,\ldots,b \\ k & = 1,2,\ldots,n \end{cases} \tag{32}$$

NHIỀU HƠN MỘT QUAN TRẮC TRONG MỘT Ô

Bảng dữ liệu cho mô hình:

Data Arrangement for a Two-Factor Factorial Design

		Factor *B* 1	2	⋯	*b*	Totals	Averages
Factor *A*	1	$y_{111}, y_{112}, \ldots, y_{11n}$	$y_{121}, y_{122}, \ldots, y_{12n}$		$y_{1b1}, y_{1b2}, \ldots, y_{1bn}$	$y_{1..}$	$\bar{y}_{1..}$
	2	$y_{211}, y_{212}, \ldots, y_{21n}$	$y_{221}, y_{222}, \ldots, y_{22n}$		$y_{2b1}, y_{2b2}, \ldots, y_{2bn}$	$y_{2..}$	$\bar{y}_{2..}$
	⋮						
	a	$y_{a11}, y_{a12}, \ldots, y_{a1n}$	$y_{a21}, y_{a22}, \ldots, y_{a2n}$		$y_{ab1}, y_{ab2}, \ldots, y_{abn}$	$y_{a..}$	$\bar{y}_{a..}$
Totals		$y_{.1.}$	$y_{.2.}$		$y_{.b.}$	$y_{...}$	
Averages		$\bar{y}_{.1.}$	$\bar{y}_{.2.}$		$\bar{y}_{.b.}$		$\bar{y}_{...}$

PHÂN CHIA SỰ BIẾN THIÊN

SST = SSG + SSB + SSI + SSE

Bậc tự do:

SST
Sự biến thiên toàn phần

abn – 1

SSG
Sự biến thiên giữa các nhóm

a – 1

SSB
Sự biến thiên giữa các khối

b – 1

SSI
Sự biến thiên do tương tác giữa nhóm và khối

(a – 1)(b – 1)

SSE
Sự biến thiên ngẫu nhiên (sai số)

ab(n – 1)

Các ký hiệu

Ta định nghĩa các đại lượng sau:

$$y_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}; \quad \bar{y}_{i..} = \frac{y_{i..}}{bn} \quad i = 1,2,\ldots,a$$

$$y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk}; \quad \bar{y}_{.j.} = \frac{y_{.j.}}{an} \quad j = 1,2,\ldots,b$$

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk}; \quad \bar{y}_{ij.} = \frac{y_{ij.}}{n} \quad \begin{cases} i & = 1,2,\ldots,a \\ j & = 1,2,\ldots,b \end{cases}$$

$$y_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}; \quad \bar{y}_{...} = \frac{y_{...}}{abn}$$

TỔNG CÁC BÌNH PHƯƠNG VỚI TƯƠNG TÁC

$$\text{Toàn phần}: SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2 \tag{33}$$

$$\text{Giữa các nhóm}: SSG = bn\sum_{j=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 \tag{34}$$

$$\text{Giữa các khối}: SSB = an\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2 \tag{35}$$

$$\text{Tương tác}: SSI = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \tag{36}$$

Trung bình bình phương

Các trung bình bình phương được định nghĩa như sau:

$$MST = \frac{SST}{abn-1} \tag{37}$$

$$MSG = \frac{SSG}{a-1} \tag{38}$$

$$MSB = \frac{SSB}{b-1} \tag{39}$$

$$MSI = \frac{SSI}{(a-1)(b-1)} \tag{40}$$

$$MSE = \frac{SSE}{ab(n-1)} \tag{41}$$

GIẢ THUYẾT CỦA BÀI TOÁN ANOVA 2 NHÂN TỐ ($n > 1$)

- Đối với các nhóm (nhân tố A):

$$\begin{cases} H_{0a}: \tau_1 = \tau_2 = \ldots = \tau_a = 0 \\ H_{1a}: \tau_i \neq 0 \text{ với ít nhất một } i \end{cases} \quad (42)$$

  - Thống kê kiểm định F:

$$F_a = \frac{MSG}{MSE} \quad (43)$$

  - Bác bỏ H_{0a} khi:

$$f_a > f_{\alpha; a-1, ab(n-1)}$$

GIẢ THUYẾT CỦA BÀI TOÁN ANOVA 2 NHÂN TỐ ($n > 1$)

- Đối với các khối (nhân tố B):

$$\begin{cases} H_{0b}: \beta_1 = \beta_2 = \ldots = \beta_b = 0 \\ H_{1b}: \beta_j \neq 0 \text{ với ít nhất một } j \end{cases} \tag{44}$$

  - Thống kê kiểm định F:

$$F_b = \frac{MSB}{MSE} \tag{45}$$

  - Bác bỏ H_{0b} khi:

$$f_b > f_{\alpha;b-1,ab(n-1)}$$

GIẢ THUYẾT CỦA BÀI TOÁN ANOVA 2 NHÂN TỐ $(n > 1)$

- Đối với tương tác giữa A và B

$$\begin{cases} H_{0b} : (\tau\beta)_{11} = (\tau\beta)_{12} = \ldots = (\tau\beta)_{ab} = 0 \\ H_{1b} : (\tau\beta)_{ij} \neq 0 \text{ với ít nhất một cặp } (i, j) \end{cases} \tag{46}$$

- Thống kê kiểm định F:

$$F_{ab} = \frac{MSI}{MSE} \tag{47}$$

- Bác bỏ H_{0ab} khi:

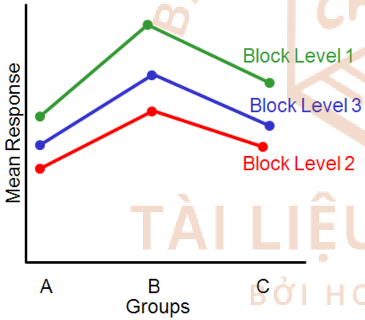
$$f_{ab} > f_{\alpha;(a-1)(b-1),ab(n-1)}$$

BẢNG ANOVA 2 NHÂN TỐ ($n > 1$)

Nguồn của sự biến thiên	SS	df	MS	F
Giữa các nhóm	SSG	$a-1$	MSG	$F_a = \frac{MSG}{MSE}$
Giữa các khối	SSB	$b-1$	MSB	$F_b = \frac{MSB}{MSE}$
Tương tác	SSI	$(a-1)(b-1)$	MSI	$F_{ab} = \frac{MSI}{MSE}$
Sai số	SSE	$ab(n-1)$	MSE	
Tổng	SST	$abn-1$		

VÍ DỤ VỀ TƯƠNG TÁC

- Không có tương tác:



- Có tương tác: