Chương 3

SO SÁNH CÁC THAM SỐ

NỘI DUNG

So sánh hai trung bình và mở rộng

- Phương pháp tham số
- Phương pháp phi tham số

So sánh hai phương sai và mở rộng

- Cơ sở lý luận
- So sánh hai phương sai
- Đánh giá sự đồng nhất các phương sai của nhiều tổng thể

Đánh giá tính độc lập của các dấu hiệu định tính

SO SÁNH HAI TRUNG BÌNH VÀ MỞ RỘNG

PHƯƠNG PHÁP THAM SỐ

Cơ sở lý luận Công thức xác định khoảng khác biệt tối thiểu có ý nghĩa phân biệt giữa chúng (**L**east **S**ignificant **D**ifference - LSD)

$$LSD_{\alpha} = t_{\alpha} Sd$$

t_{α} là giá trị tới hạn phân phối Student ở mức α

Sd là sai số thực nghiệm giữa hai trung bình

Ở độ tin cậy 1 - α khi $|X_1 - X_2| < LSD_{\alpha}$

=> $X_1 = X_2$ và ngược lại

Để thuận tiện trong cách diễn đạt người ta lập "giả thuyết"

$$H_0 : X_1 = X_2 \;;\; H_1 : X_1 \neq X_2$$

=> Chấp nhận giả thuyết H_0 hoặc từ chối giả thuyết H_0

Tuy nhiên do

$$T_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{Sd} \quad \text{và} \quad t_\alpha = \frac{LSD_\alpha}{Sd}$$

Nên thay vì kiểm định sự chênh lệch giữa hai trung bình $|X_1 - X_2|$ so với LSD_α, người ta chuyển sang kiểm định T_{TN} so với $t_{bảng}$.

Khi $T_{TN} < t_{bảng}$ => giả thuyết H_0 được chấp nhận

Khi $T_{TN} > t_{bảng}$ giả thuyết H_1 được chấp nhận.

Trong trường hợp dung lượng mẫu lớn hoặc đã biết phương sai của hai tổng thể thì có thể tính U_{TN}

$$U_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{Sd}$$ được so với $u_{\alpha/2}$.

So sánh hai trung bình khi đã biết phương sai của hai tổng thể σ_1^2 và σ_2^2

Công thức tính U_{TN}

$$U_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad Sd = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- X_1 và X_2 là trung bình của hai mẫu mẫu quan sát
- σ_1^2 và σ_2^2 là phương sai của hai mẫu quan sát
- n_1 và n_2 là dung lượng của hai mẫu quan sát; Sd lúc này được tính
- Nếu $U_{TN} < u_{\alpha/2}$ thì chấp nhận giả thuyết H_0 ở độ tin cậy $1 - \alpha$.
- Nếu $U_{TN} > u_{\alpha/2}$ thì chấp nhận giả thuyết H_1 ở độ tin cậy $1 - \alpha$.

So sánh hai trung bình khi chưa biết phương sai nhưng biết chúng bằng nhau ($\sigma_1^2 = \sigma_2^2$)

Tính phương sai mẫu và kiểm tra S_1^2 và S_2^2 nhờ phép trắc nghiệm F

$$F_{TN} = \frac{S_1^2}{S_2^2}$$

Nếu $F_{TN} < F_{bảng}$ => hai phương sai bằng nhau và ngược lại

Khi $S_1^2 = S_2^2$, thì việc so sánh giữa hai trung bình được thực hiện theo công thức

$$T_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\text{ở đây } Sd = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

t_α được tra với độ tự do $(n_1 + n_2 - 2)$

Ví dụ: So sánh năng suất cá thể của tổ hợp bông lai F_1 C92-52/C118A theo số liệu sau đây và F_1 S02-13/TM1 (ở ví dụ Bảng 2.1, chương 2) với các thông tin sau:

Tổ hợp C92-52/C118A: $n_1 = 45$; $\overline{x} = 63{,}56$, $S^2 = 387{,}90$

Tổ hợp S02-13/TM1: $n_2 = 50$; $\overline{x} = 76{,}92$, $S^2{}_1 = 351{,}68$

Giải:

$$F_{TN} = \frac{387{,}90}{351{,}68} = 1{,}10$$

Tra bảng F với hai độ tự do 49 và 44 ta có $F_{0,05} = 1{,}63$ => Như vậy hai phương sai bằng nhau

$$T_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \frac{76,92 - 63,56}{\sqrt{\frac{(50-1)(351,68)+(45-1)(387,90)}{50+45-2}\left(\frac{1}{50}+\frac{1}{45}\right)}}$$

$= 3.38$

Tra t_α với độ tự do $(50 + 45 - 2) = 93$ ta được:

$t^{93}{}_{0.05} = 1.99$, $t^{93}{}_{0.01} = 2.63$.

$T_{TN} = 3,38 > t^{93}{}_{0.01} = 2,63$

$\Rightarrow$ Năng suất F1 tổ hợp S02-13/TM1 cao hơn tổ hợp C92-52/C118A với độ tin cậy 99%.

Kết quả so sánh trung bình F1 S02-13/TM1 và F1 C92-52/C118A trên phần mềm Excel (lưu ý thí dụ này không có số liệu thô)

t-Test: Two-Sample Assuming Equal Variances

	C92-52/C118A	S02-13/TM1
Mean	63.56	76.922
Variance	387.90	351.68
Observations	45	50
Pooled Variance	368.8169	
Hypothesized Mean Difference	0	
df	93	
t Stat	-3.3861	
P(T<=t) one-tail	0.0005	
t Critical one-tail	1.6614	
P(T<=t) two-tail	0.0010	
t Critical two-tail	1.9858	

So sánh hai trung bình khi chưa biết phương sai nhưng biết rằng chúng khác nhau ($\sigma_1^2 \neq \sigma_2^2$)

Khi n > 30 và khi $\boldsymbol{\sigma_1^2 \neq \sigma_2^2}$, việc so sánh giữa hai trung bình được thực hiện theo công thức

$$T_{TN} = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

- X_1 và X_2 là trung bình của hai mẫu mẫu quan sát;
- σ_1^2 và σ_2^2 là phương sai của hai mẫu quan sát;
- n_1 và n_2 là dung lượng của hai mẫu quan sát

Giá trị t_α được tra với **<u>k độ tự do</u>** lấy số nguyên từ công thức sau

$$k = \frac{(n_1 - 1)(n_2 - 1)\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{(n_1 - 1)\left(\frac{S_1^2}{n_1}\right)^2 + (n_2 - 1)\left(\frac{S_2^2}{n_2}\right)^2}$$

- Nếu $T_{TN} < t_{bảng}$ ở mức α thì kết luận rằng $X_1 = X_2$ ở độ tin cậy $1 - \alpha$
- Nếu $T_{TN} > t_{bảng}$ ở mức α thì kết luận rằng $X_1 \neq X_2$ ở độ tin cậy $1 - \alpha$.

Ví dụ: So sánh năng suất cá thể thế hệ F_1 và F_2 của tổ hợp bông lai C92-52/C118A trồng tại Đại học Nông Lâm Tp. HCM 2008 theo kết quả theo dõi sau đây.

Năng suất cá thể (g/cây) của 45 cây F_1

50,7	30,0	32,9	78,1	41,3	72,9	57,1	52,0	94,5	87,7
69,7	64,6	72,9	79,6	91,2	46,6	42,9	42,9	29,4	76,4
72,0	65,8	58,1	50,1	53,1	71,0	54,5	52,1	62,3	94,0
59,2	38,5	57,9	66,0	39,6	78,6	37,4	54,8	78,4	48,6
98,0	68,0	96,8	97,8	94,2					

Trung bình: 63,56; Phương sai: 387,90; Độ lệch chuẩn:19,70

Năng suất cá thể (g/cây) của 110 cây F_2

20,9	69,4	42,5	21,3	45,6	21,5	14,9	10,7	11,3	20,3
103,0	10,4	97,0	53,0	57,5	41,8	79,5	91,0	44,5	37,0
42,0	11,7	54,9	41,8	49,2	52,4	55,1	91,0	47,5	43,0
49,6	64,3	132,0	60,7	94,0	4,5	99,0	96,3	89,4	96,0
49,5	59,1	44,9	42,9	62,8	49,7	73,8	46,9	75,8	62,0
40,2	57,9	87,7	53,3	98,5	3,2	98,2	41,9	58,8	79,1
49,5	52,3	63,8	17,4	77,6	69,9	65,5	59,6	79,5	48,5
17,7	38,0	20,5	35,9	47,4	37,0	85,8	45,5	29,0	62,8
28,9	31,6	16,6	34,6	48,5	37,4	64,2	50,4	26,5	94,0
78,0	16,6	37,8	38,1	83,3	86,4	29,6	25,5	33,4	11,3
74,2	19,9	75,8	59,1	33,1	66,4	139	52,4	31,8	98,0

Trung bình: 53,45; Phương sai: 768,61; Độ lệch chuẩn: 27,72

$$T_{TN} = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{63{,}56 - 53{,}45}{\sqrt{\frac{387{,}9}{45} + \frac{768{,}61}{110}}} = 2{,}56$$

- Thay các giá trị vào công thức tính độ tự do k, ta có k = 136
- Với độ tự do này tiêu chuẩn T ≅ tiêu chuẩn U
- Do đó $t^{136}{}_{0.05} \cong u_{0.025} = 1{,}98$
  còn $t^{136}{}_{0.01} \cong u_{0.005} = 2{,}61$.
- Như vậy, năng suất F1 cao hơn năng suất F2 với độ tin cậy trên 95% gần 99%.

Kết quả so sánh trung bình F1 và F2 trên phần mềm Excel:

t-Test: Two-Sample Assuming Unequal Variances

	F1	F2
Mean	63.56	53.45
Variance	387.90	768.61
Observations	45	110
Hypothesized Mean Difference	0	
df	114	
t Stat	2.5596	
P(T<=t) one-tail	0.0059	
t Critical one-tail	1.6583	
P(T<=t) two-tail	0.0118	
t Critical two-tail	1.9810	

- Khi n < 30, và khi hai phương sai mẫu $S_1^2 \neq S_2^2$ việc so sánh sẽ kém chính xác.

- Trong trường hợp này có thể áp dụng phương pháp rút mẫu ngẫu nhiên có hoàn lại từ mẫu đã có rất nhiều lần để ước lượng trung bình mới của hai mẫu và tiến hành so sánh như trường hợp dung lượng mẫu lớn

So sánh hai trung bình lấy mẫu theo cặp (Paired two samples)

Công thức tính T_{TN}

$$T_{TN} = \frac{\overline{d}}{S_d}\sqrt{n}$$

trong đó:

$\overline{d} = \frac{1}{n}\sum_{i=1}^{n}(x_{1i} - x_{2i})$ là trung bình độ lệch;

$S_d = \sqrt{V(d_i)}$ là độ lệch chuẩn của các độ lệch $d_i = x_{1i} - x_{2i}$

n là dung lượng mẫu ($n = n_{x_1} = n_{x_2}$)

Nếu $T_{TN} \geq t_{0,05}$ với n – 1 bậc tự do thì $\overline{X}_1 \neq \overline{X}_2$, ngược lại $T_{TN} < t_{0,05}$ thì $\overline{X}_1 \approx \overline{X}_2$.

Ví dụ: Kết quả học tập của 26 sinh viên năm thứ nhất và năm thứ 2 được ghi ở Bảng 3.1.

Bảng 3.1: Điểm trung bình các môn của 26 sinh viên

TT	Điểm các năm		$X_2 - X_1$	TT	Điểm các năm		$X_2 - X_1$
	1 (X_1)	2(X_2)	(d_i)		1 (X_1)	2(X_2)	(d_i)
1	8,3	7,9	– 0,4	14	6,7	7,2	+ 0,5
2	8,4	8,6	+ 0,2	15	6,9	7,1	+ 0,2
3	8,2	8,1	– 0,1	16	9,3	9,4	+ 0,1
4	6,5	7,2	+ 0,7	17	5,8	5,6	– 0,2
5	7,8	7,3	– 0,5	18	8,6	8,8	+ 0,2
6	6,9	7,2	+ 0,3	19	6,2	5,8	– 0,4
7	7,1	7,1	0,0	20	7,9	8,2	+ 0,3
8	8,4	8,7	+ 0,3	21	7,8	7,6	– 0,2
9	7,6	7,9	+ 0,3	22	9,0	8,7	– 0,3
10	7,8	7,7	– 0,1	23	8,2	8,6	+ 0,4
11	7,5	7,7	+ 0,2	24	8,4	8,2	– 0,2
12	6,4	7,2	+ 0,8	25	7,3	7,2	– 0,1
13	6,8	7,1	+ 0,3	26	5,8	5,4	– 0,4

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n}(x_{1i} - x_{2i}) = (-\,0{,}4 + 0{,}2 - 0{,}1 + \ldots$$

$$- 0{,}4) = 0{,}073$$

$$S_d = \sqrt{V(d_i)} = 0{,}346$$

Cuối cùng: $T_{TN} = \frac{\overline{d}}{S_d}\sqrt{n} = \frac{0{,}073}{0{,}346}\sqrt{26} = 1{,}08$

$$T_{TN} = 1{,}08 < t_{0,05}^{(26-1)} = 2{,}06.$$

Như vậy kết quả học tập hai năm như nhau ($\overline{X}_1 \approx \overline{X}_2$).

PHƯƠNG PHÁP PHI THAM SỐ

- Với phương pháp phi tham số, các tiêu chuẩn kiểm định dựa vào **thứ hạng xếp** theo độ lớn nhỏ của các giá trị quan sát, không sử dụng tham số trung bình và phương sai.

- Phương pháp phi tham số không chính xác bằng các phương pháp tham số

So sánh các trung bình các mẫu độc lập

• So sánh trung bình hai mẫu độc lập

Ví dụ: Để đánh giá tính đồng nhất của khu đất thí nghiệm tại Trại thực nghiệm Đại học Nông Lâm Tp. HCM, chúng tôi đã tiến hành đo chiều cao cây của giống bông S02-13 được trồng ở ba vị trí (ba lô), mỗi lô theo dõi chiều cao 10 cây. Kết quả như sau:

Chiều cao cây (cm) của các cây trong ba lô

Lô 1:	72	87	71	70	80	67	80	80	82	66
Lô 2:	97	95	90	81	92	91	95	96	84	72
Lô 3:	62	68	73	85	69	79	77	76	83	84

Bước 1: Xếp hạng số liệu – kết quả như sau

Lô 1	72	87	71	70	80	67	80	80	82	66
Hạng	*(5)*	*13*	*4*	*3*	*(7)*	*2*	*(7)*	*(7)*	*11*	*1*
Lô 2	97	95	90	81	92	91	95	96	84	72
Hạng	*20*	*17*	*14*	*10*	*16*	*15*	*17*	*19*	*12*	*(5)*

Ở đây có 2 hạng 5 cho số 72 theo thứ tự 5, 6; 3 hạng 7 cho số 80 theo thứ tự 7, 8, 9, vì thế mỗi số 72 có thứ hạng mới là 5,5, tức là (5 + 6)/2 và mỗi số 80 có thứ hạng mới là 8, tức là (7 + 8 + 9)/3. Việc xếp hạng đúng khi:

$$R_1 + R_2 = \frac{n(n+1)}{2}$$

R_1 là tổng thứ hạng của lô 1 và R_2 là tổng thứ hạng của lô 2.

Bước 2: Kiểm tra và đánh giá kết quả

$$U_{TN} = \frac{\left| U_1 - \frac{n_1 n_2}{2} \right|}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

Nếu $U_{TN} > 1,96$ thì $U_1 \neq U_2$; ngược lại

$U_{TN} < 1,96$ thì $U_1 \cong U_2$.

Ở ví dụ này: R1 = 63,5; R2 = 146,5. Thay vào công thức ta có: U1 = 91,5 và U2 = 8,5.

$$U_{TN} = \frac{\left|91,5 - \frac{10\times10}{2}\right|}{\sqrt{\frac{10\times10(10+10+1)}{12}}} = 3,14 > 1,96$$

Tương tự, kết quả kiểm tra U_1 (lô 1) và U_3 (lô 3) ta được

$R_1 = 104$; $R_3 = 106$;

$U_1 = 51{,}0$; $U_3 = 49{,}0$;

$U_{TN} = 0{,}08$

Giữa U_2 (lô 2) và U_3 (lô 3):

$R_2 = 143{,}5$; $R_3 = 66{,}5$;

$U_1 = 11{,}5$; $U_3 = 88{,}5$;

$U_{TN} = 2{,}91$

Với các kết quả này thì đất lô 1 và lô 3 đồng nhất và khác với lô 2 về độ phì nhieu

So sánh các trung bình nhiều mẫu độc lập

Công thức tính H

$$H = \frac{12}{n(n+1)} \sum_{k=1}^{k} \frac{R_i^2}{n_i} - 3(n+1) \qquad (3 -$$

Trong đó: $n = \Sigma n_i$ là tổng dung lượng mẫu

R_i là tổng các hạng trong mẫu i $(i = \overline{1,k})$

k là số mẫu quan sát

Nếu H > $\chi^2_{0.05}$ thì các mẫu không thuần nhất.

Nếu H < $\chi^2_{0.05}$ thì các mẫu thuần nhất

Ví dụ: Từ kết quả phân tích hàm lượng mùn (%) trong 3 lô thí nghiệm ở bảng ở trang sau, hãy so sánh tính đồng nhất của khu đất.

Lô 1:	12,3	12,5	13,1	13,6	13,8	14,2	14,7	14,9	15,3	n_1=9
Lô 2:	12,8	13,9	14,2	14,7	15,3	15,3	15,8	16,8	17,4	n_2=9
Lô 3:	13,9	14,9	15,7	15,7	15,8	16,5	16,8	17,3	18,5	n_3=9

Áp dụng phương pháp xếp hạng như trên ta có:

Lô 1:	12,3	12,5	13,1	13,6	13,8	14,2	14,7	14,9	15,3	n_1=9
Hạng	*1*	*2*	*4*	*5*	*6*	*9,5*	*11,5*	*13,5*	*16*	R_1=68,5
Lô 2:	12,8	13,9	14,2	14,7	15,3	15,3	15,8	16,8	17,4	n_2=9
Hạng	*3*	*7,5*	*9,5*	*11,5*	*16*	*16*	*20,5*	*23*	*26*	R_2=133
Lô 3:	13,9	14,9	15,7	15,7	15,8	16,5	16,8	17,3	18,5	n_3=9
Hạng	*7,5*	*13,5*	*18,5*	*18,5*	*20,5*	*22*	*24*	*25*	*27*	R_3=176,5

$$\sum_{i=1}^{k} R_i = 68,5 + 133 + 186,5 = 378$$

Kiểm tra lại kết quả xếp hạng:

$$\sum_{i=1}^{k} R_i = \frac{27(27+1)}{2} = 378$$

Như vậy việc xếp hạng là đúng.

Theo công thức trên ta có: $H = 10,34$, $\chi^2_{0,05} = 5,99$.

Như vậy, không có sự đồng nhất của 3 lô đất thí nghiệm.

So sánh trung bình hai mẫu phụ thuộc

Nếu các tổng thể lại không theo luật phân phối chuẩn thì việc so sánh được thực hiện bằng phép nghiệm phi tham số Wilcoxon

Các bước thực hiện

1. Xếp hạng từ nhỏ đến lớn các số đo của cả hai mẫu.

2. Tính kỳ vọng và phương sai

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma_T^2 = \frac{n_1 n_2}{12}\left[n_1 + n_2 + 1\right] - \frac{\sum_j t_j\left(t_j^2 - 1\right)}{\left(n_1 + n_2\right)\left(n_1 + n_2 - 1\right)}$$

Nếu n_1 và $n_2 \leq 10$

Sau khi tính được tổng hạng của mỗi mẫu, tra bảng giá trị tổng hạng Wilcoxon để tìm các giá trị tới hạn T_L và T_u và xác định:

- Nếu kỳ vọng của hai tổng thể giống nhau thì $T < T_u$ (hoặc $T > T_L$).

- Nếu kỳ vọng của hai tổng thể khác nhau thì $T > T_u$ (hoặc $T < T_L$).

Nếu n_1 và n_2 > 10

Trong kiểm định tổng hạng Wilcoxon khi cả n_1 và n_2 đều lớn hơn 10 thì phân phối T sẽ tiệm cận với phân phối chuẩn U. Khi đó việc so sánh trung bình của hai mẫu theo tiêu chuẩn U

$$U_{TN} = \frac{T - \mu_T}{\sigma_T}$$

Nếu $U_{TN} > u_{\alpha/2}$ thì X_1 khác X_2 ở độ tin cậy $1 - \alpha$.

Nếu $U_{TN} < u_{\alpha/2}$ thì X_1 không khác với X_2 ở độ tin cậy $1 - \alpha$.

Ví dụ: Để đánh giá một giống bắp mới trong sản xuất một công ty đã hợp đồng với 15 gia đình trong vùng, yêu cầu mỗi gia đình chọn một đám đất đồng đều và gieo 2 giống: giống đang sản xuất phổ biến trong vùng làm đối chứng (ĐC) và giống mới do công ty mang xuống (GM). Họ còn yêu cầu việc thực hiện các công việc gieo trồng, chăm sóc phải thực hiện như nhau cho cả hai giống. Kết quả năng suất và phân hạng (trong ngoặc) được ghi trong Bảng 3.2 sau đây. Hãy so sánh năng suất 2 giống này ?

Bảng 3.2: Năng suất và phân hạng hai giống bắp

Gia đình	Năng suất (tạ/ha)		Gia đình	Năng suất (tạ/ha)	
	ĐC	GM		ĐC	GM
1	68,5 (*5,5*)	77,5 (*18*)	9	69,0 (*7*)	71,8 (*12*)
2	73,3 (*15*)	83,0 (*22*)	10	72,0 (*13*)	74,8 (*16*)
3	57,5 (*1*)	62,5 (*2*)	11	79,0 (*20*)	88,0 (*29*)
4	81,4 (*21*)	87,6 (*28*)	12	69,5 (*9*)	72,2 (*14*)
5	77,8 (*19*)	69,3 (*8*)	13	87,3 (*27*)	83,8 (*24*)
6	85,5 (*25*)	93,5 (*30*)	14	70,2 (*10*)	63,8 (*3*)
7	68,5 (*5,5*)	71,5 (*11*)	15	68,0 (*4*)	76,5 (*17*)
8	87,2 (*26*)	83,5 (*23*)		T = 465	

Ta có : $n_1 = n_2 = 15$; $n = n_1 + n_2 = 30$; $T = 465$

Tính kỳ vọng của tổng thể:

$$\mu_T = (15 \times 31)/2 = 232{,}5$$

Tính phương sai σ_T^2 và U_{TN}

$$\sigma_T^2 = \frac{225}{12} \times 31 - \frac{6}{30 \times 29} = 581{,}24 \text{ và } \sigma_T = 24{,}11$$

$$\text{Từ đó: } U_{TN} = \frac{T - \mu_T}{\sigma_T} = \frac{465 - 232{,}5}{24{,}11} = 9{,}64$$

Với $\alpha = 0{,}05$, $u_{0,025} = 1{,}96$ và $\alpha = 0{,}01$, $u_{0,005} = 2{,}58$

Như vậy: giống GM > ĐC với độ tin cậy 99%.

So sánh các trung bình nhiều mẫu phụ thuộc

Việc so sánh được thực hiện bằng phép thử Friedman.

Các bước thực hiện

- Xếp hạng thứ tự 1, 2, 3, ... giữa các phương án trong từng nơi (hoặc từng thời điểm), mỗi nơi một hàng.
- Tính tổng số hạng cho từng phương án theo từng cột.
- Kiểm tra sự giống hay khác nhau giữa các phương án theo tiêu chuẩn χ^2

$$\chi^2_{TN} = \frac{12}{ab(a+1)}\sum R_i^2 - 3b(a+1)$$

Trong đó: a là số phương án

b là số nơi (số thời điểm)

R_i là tổng hạng của từng phương án $(i = \overline{1,a})$.

Nếu $\chi_{TN}^2 < \chi_{0.05}^2$ thì các phương án khác nhau không đủ tin cậy.

Nếu $\chi_{TN}^2 > \chi_{0.05}^2$ với 1 - α độ tự do thì các phương án cho kết quả khác nhau

Ví dụ 1: Trong một thử nghiệm 5 giống đậu xanh tại 3 xã Phước Tiến, Phước Thắng và Phước Đại, huyện Bác Ái, tỉnh Ninh Thuận vụ Hè Thu năm 2009, năng suất các giống tại các điểm thí nghiệm được ghi nhận trong Bảng 3.3. Câu hỏi đặt ra: năng suất của các xã này khác nhau không ?

Bảng 3.3: Năng suất (tạ/ha) của 5 giống đậu xanh và kết quả xếp hạng từng xã cho mỗi giống

Giống\Xã	Phước Tiến /hạng		Phước Thắng /hạng		Phước Đại /hạng	
NP 305	13,9 b	*3*	12,3 c	*1*	13,8 bc	*2*
ĐX208	17,8 a	*3*	16,4 a	2	15,6 ab	*1*
HL 89-E3	16,2 a	*3*	15,1 b	*1*	16,1 a	*2*
V 99-1	12,3 b	*2*	12,2 c	*1*	12,5 c	*3*
Agredec-01	13,6 b	*2*	12,9 c	*1*	14,6 ab	*3*
P	< 0,05		< 0,05		< 0,05	
ΣR_i		13		6		11

Ta có : a = 3, b = 5,

$SR_1 = 13$, $SR_2 = 6$, $SR_3 = 11$

Tính χ_{TN}^2:

$$\chi_{TN}^2 = \frac{12}{3x5(3+1)}\left(13^2+6^2+11^2\right)-3x5x4$$

$$= 5{,}20 < \chi_{0.05}^{2(2)} = 6{,}0.$$

Như vậy năng suất đậu xanh của 3 xã này không có sự khác nhau

Ví dụ 2: Kết quả cân khối lượng 100 cây mầm một giống cải bẹ xanh *Brassica juncea* L. được gieo trên 4 loại giá thể khác nhau (TN1, NT2, NT3 và TN4) tại Bảo Lộc, Lâm Đồng được ghi ở bảng 3.4. Theo kết quả này, có sự khác nhau hay không về khối lượng cây mầm trên các loại giá thể?

Bảng 3.4: Khối lượng trung bình 100 cây mầm (g) của 4 nghiệm thức (NT) trên các lần lặp lại.

Lặp lại	Khối lượng trung bình 100 cây mầm (g)			
	NT1	NT2	NT3	NT4
I	4,5 (*1*)	5,5 (*4*)	5,2 (*2,5*)	5,2 (*2,5*)
II	4,7 (*1*)	5,3 (*4*)	4,8 (*2*)	5,0 (*3*)
III	4,7 (*1*)	5,1 (*4*)	5,0 (*3*)	4,8 (*2*)
ΣR_i	3	12	7,5	7,5

SO SÁNH HAI PHƯƠNG SAI VÀ MỞ RỘNG

So sánh hai phương sai

$$F_{TN} = \frac{S_1^2}{S_2^2} \qquad (S_1^2 > S_2^2)$$

* Nếu FTN < f_α thì $S_1^2 \cong S_2^2$

* Nếu FTN > f_α thì $S_1^2 > S_2^2$ ở độ tin cậy 1 - α

- **Đánh giá sự đồng nhất các phương sai của nhiều tổng thể**
  - **Khi dung lượng mẫu rút ra từ các tổng thể khác nhau**
    - Nếu dung lượng mẫu của k phương sai mẫu S_1^2. S_2^2, ...S_k^2 là n1, n2, ... nk (i =1, k)
    - $n_1 \neq n_2 \neq \ldots \neq n_k$; $h_i = (n_i - 1)$, $h = \Sigma h_i$
    - S^2 là trung bình số học của k phương sai

$$\overline{S}^2 = \frac{\sum h_i S_i^2}{h}$$

Để kiểm định sự đồng nhất của các phương sai ta có

$$B = \frac{V}{C}, \text{ trong đó:}$$

$$V = 2,303\left[h \lg \overline{S}^2 - \sum h_i \lg S_i^2\right]$$

$$C = 1 + \frac{1}{3(k-1)}\left[\sum \frac{1}{h_i} - \frac{1}{h}\right]$$

$B < \chi^{2(k-1)}{}_{0.05}$ => *c*ác phương sai đồng nhất

$B \geq \chi^{2(k-1)}{}_{0.05}$ => các phương sai không đồng nhất

Ví dụ: Kết quả điều tra biến động năng suất cá thể (g/cây) của 5 giống bông thuần với $n_1 = 26$, $n_2 = 32$, $n_3 = 29$, $n_4 = 30$, $n_5 = 19$ và các phương sai tương ứng $S_1^2 = 623{,}8$, $S_2^2 = 420{,}4$, $S_3^2 = 630{,}6$, $S_4^2 = 461{,}0$ và $S_5^2 = 586{,}6$. Hãy kiểm định tính đồng nhất của các phương sai với độ tin cậy 95%.

Để tính B ta lập bảng sau

Mẫu	h_i	S_i^2	$h_i S_i^2$	$\lg S_i^2$	$h_i \lg S_i^2$	$1/h_i$
1	25	623,8	15.595,0	2,795	69,9	0,040
2	31	420,4	13.032,4	2,624	81,3	0,032
3	28	630,6	17.656,8	2,800	78,4	0,036
4	29	461,0	13.369,0	2,664	77,2	0,034
5	18	586,6	10.558,8	2,768	49,8	0,056
Σ	131		70.212,0		356,7	0,198

Từ đó:

$$\overline{S}^2 = \frac{\sum h_i S_i^2}{h} = \frac{70.212,0}{131} = 536,0$$

$$\lg \overline{S}^2 = 2,729$$

$$V = 2,303\left[h \lg \overline{S}^2 - \sum h_i \lg S_i^2\right]$$

$$= 2,303[(131)(2,729) - 356,7] = 1,982$$

$$C = 1 + \frac{1}{3(k-1)}\left[\sum \frac{1}{h_i} - \frac{1}{h}\right]$$

$$= 1 + \frac{1}{3(5-1)}\left[0,198 - \frac{1}{131}\right] = 1,016$$

$$B = \frac{V}{C} = \frac{1,982}{1,016} = 1,898 < \chi^{2(4)}_{0,05} = 9,488$$

Kết luận: các phương sai được xem là đồng nhất, tức là các giống đều thuần chủng

- **Khi dung lượng mẫu rút ra từ các tổng thể bằng nhau**

$$G = \frac{S_{max}^2}{S_1^2 + S_2^2 + \ldots + S_k^2}$$

$G < g_\alpha^{(n-1,k)}$ => các phương sai mẫu đồng nhất

$G \geq g_\alpha^{(n-1,k)}$ => ác phương sai không đồng nhất

Ví dụ: Cũng với các giống trên, nếu lấy dung lượng mẫu bằng nhau và bằng 19, các phương sai mẫu sẽ là 653,2; 466,1; 671,3; 581,4 và 586,6. Hãy kiểm định tính đồng nhất của các phương sai với độ tin cậy 95% và ước lượng phương sai tổng thể nếu các phương sai đồng nhất.

Giải

$$G = \frac{671{,}3}{653{,}2+466{,}1+671{,}3+581{,}4+586{,}6} = 0{,}227$$

Với $\alpha = 0,05$; số bậc tự do là $19 - 1 = 18$; và số lượng mẫu là 5

=> giá trị tới hạn tra được là $g_a^{(n-1,k)} =$

$$g_{0.05}^{(18,5)} = 0,3645$$

$G < g_{0.05}^{(18,5)}$ cho thấy các phương sai là đồng nhất, và phương sai tổng thể được ước lượng

$$\sigma^2 = \overline{S}^2 = \frac{653,2 + 466,1 + 671,3 + 581,4 + 586,6}{5}$$

$$= 591,7$$

ĐÁNH GIÁ TÍNH ĐỘC LẬP CỦA CÁC DẤU HIỆU ĐỊNH TÍNH

Người ta sử dụng trắc nghiệm CHI bình phương (χ^2) để xác định mối quan hệ giữa hai dấu hiệu định tính

Để kiểm tra các giả thiết này, từ tổng thể có dung lượng mẫu **n**, lập bảng trình bày các đặc trưng **A, B** và tần số tương ứng

A \ B	B_1	B_2	...	B_j	Tổng A_i
A_1	n_{11}	n_{12}	...	n_{1j}	$n_{1.}$
A_2	n_{21}	n_{22}	...	n_{2j}	$n_{2.}$
...	...	...	...	...	...
A_i	n_{i1}	n_{i2}	...	n_{ij}	$n_{i.}$
Tổng B_j	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	$\Sigma n_{ij} = n$

n là dung lượng mẫu.

$\mathbf{n_{ij}}$ là là tần số ứng với các mức độ của A_i (i =1, i) và B_j (j =1, j) .

$\mathbf{n_{i}}$. là là tần số ứng với các mức độ của dấu hiệu A.

$\mathbf{n_{j}}$ là là tần số ứng với các mức độ của dấu hiệu B.

Tính độc lập của hai dấu hiệu A và B được kiểm tra theo trắc nghiệm CHI bình phương (χ^2)

$$\chi^2_{TN} = n\left(\sum_{i=1}^{i}\sum_{j=1}^{j}\frac{n_{ij}^2}{n_{i.}n_{.j}} - 1\right)$$

Nếu ${\chi_{TN}}^2 < {\chi_\alpha}^2$ với (i – 1)(j – 1) thì chấp nhận H_0 ở độ tin cậy 1 - α.

Nếu ${\chi_{TN}}^2 \geq {\chi_\alpha}^2$ độ tự do thì chấp nhận H_1 ở độ tin cậy 1 – α

Ví dụ: Kết quả điều tra mức độ lông của lá bông và mức độ kháng rầy xanh được ghi ở Bảng 3.5. Vậy, tính có lông có quan hệ với mức độ kháng rầy không?

Bảng 3.5: Kết quả điều tra tính kháng rầy 50 giống bông

Mức độ lông của lá	KTB	K	RB	Tổng, A_i
Ít	3	0	0	3
Vừa	7	3	0	10
Nhiều	5	8	5	18
Rất nhiêu	3	5	11	19
Tổng, B_j	18	16	16	$\Sigma n_{ij} = 50$

Ghi chú : KTB – kháng trung bình; K – kháng ; RK – rất kháng.

Ở đây: i = 4; j = 3; n = 50; n_i. = 3, 10, 18 và 19; n_j = 18, 16 và 16.

Thay giá trị vào công thức ta được:

$$\chi^2_{TN} = 50 \times \{[3^2/(18 \times 3) + 7^2/(18 \times 10) + 5^2/(18 \times 18) + \ldots$$

$$+ 5^2/(16 \times 18) + 11^2/(16 \times 19)] - 1\} = 19{,}40$$

$$\chi^2_{TN} = 19{,}40 > \chi^{2(6)}_{0{,}01} = 16{,}81.$$

=> Như vậy, tính có lông có quan hệ chặt chẽ với mức độ kháng rầy với độ tin cậy 99%