

บทที่ 7 การทดสอบอัตราส่วนทางพันธุศาสตร์

1. ความหมายของ Chi-Square Test

Chi-square เป็นวิธีการหนึ่งที่ใช้วัดว่าค่าสังเกตที่ได้มานั้นกับค่าที่คาดว่าจะได้ตามทฤษฎีนั้นคล้อยตามกันหรือไม่ หรือว่าความแตกต่างระหว่างค่าที่สังเกตได้กับค่าที่ได้จากการคำนวณนั้น เกิดขึ้นจาก โอกาส (chance) หรือเกิดจากสาเหตุอื่น (เช่น สมมติฐานที่ตั้งไว้นั้นผิดไปจากความเป็นจริง)

2. การคำนวณค่า Chi-Square

ในการคำนวณค่า Chi-Square ให้นำค่า deviation หรือความแตกต่างระหว่างค่าที่สังเกตกับค่าที่คาดว่าจะได้จากการคำนวณ มายกกำลังสอง แล้วหารด้วยค่าที่คาดว่าจะได้จากการคำนวณ (expected value) นำผลลัพธ์ที่ได้ในแต่ละ class ที่เกี่ยวข้องของทุก class มารวมกันเป็นค่า Chi-Square

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

χ^2 คือค่า Chi-Square

O คือค่าที่สำรวจมาได้ (observed value)

E คือค่าที่คาดหวังว่าจะเป็น (expected value)

ยิ่งความแตกต่างระหว่าง observed value กับ expected value มีมากเท่าไร ค่า Chi-Square จะยิ่งมากขึ้นเท่านั้น นอกจากนั้นจำนวนของ class ที่เรานำค่า Chi-Square มารวมกันนั้น ก็มีส่วนที่จะทำให้ค่า Chi-Square สูงขึ้นด้วย เพราะยังมีหลาย class ก็จะทำให้ค่า Chi-Square มากขึ้น เพราะฉะนั้นจึงจำเป็นที่จะต้องพิจารณาจำนวน class เข้าไปด้วย แต่โดยเหตุว่าในการวิเคราะห์ทางสถิติกับผลการทดลองต่างๆ หรือผลที่ได้จากอัตราส่วนทางพันธุศาสตร์ เราถือว่าจำนวนค่าสังเกตที่ได้มาทั้งหมดหรือผลรวมนั้นมีค่าตายตัว เมื่อเป็นเช่นนี้บาง class ของค่าที่ได้มาก็มีความเป็นอิสระที่จะมีค่าเท่าไรก็ได้ บาง class ก็จะไม่มีความเป็นอิสระเหลืออยู่เลย ดังนั้นจึงจำเป็นที่จะต้องมิตัวเลขที่บอกถึงความเป็นอิสระของจำนวน class ขึ้น ที่เรียกว่า degree of freedom (d.f.) เช่น ถ้าเรามีถุงมืออยู่สองข้างซึ่งเป็นจำนวนตายตัวอยู่แล้ว ในการที่จะสวมถุงมือนั้น ในครั้งแรกเราอาจจะสวมถุงมือข้างใดก็ได้ (มีความเป็นอิสระในการเลือก) แต่เมื่อสวมลงไปแล้ว อีกมือหนึ่งจะต้องสวมถุงมือข้างที่เหลือ (ไม่มีความเป็นอิสระในการเลือก) และในทำนองเดียวกัน ถ้ามี phenotype อยู่หลาย class เมื่อจำนวนค่าสังเกตทั้งหมดตายตัวแล้ว class ต่างๆ อาจมีจำนวนค่าสังเกตเท่าใดก็ได้ ยกเว้น class สุดท้ายที่จะต้องมิตัวเลขตายตัว คือเป็นจำนวนเท่าที่ยังเหลืออยู่ (ไม่มีความเป็นอิสระที่จะมีจำนวนเป็นเท่าไร)

ดังนั้น degree of freedom = จำนวน class - 1

ในทางสถิติ แต่ละค่า chi-Square จะมี probability อยู่ค่าหนึ่ง ซึ่งก็ขึ้นอยู่กับค่า d.f. จึงมีการสร้างตารางสำเร็จขึ้นมาไว้เปรียบเทียบ ซึ่งหาได้จากหนังสือที่อ้างอิงไว้ข้างท้าย และได้นำบางส่วนของตารางมาแสดงเพื่อใช้ในตัวอย่างการคำนวณในเอกสารฉบับนี้แล้ว

3. การสรุปผลการทดสอบ

เมื่อคำนวณค่า Chi-Square ได้แล้วก็นำไปเปรียบเทียบกับค่า Chi-Square ในตาราง ถ้าหากค่า Chi-Square ที่ได้มา มีค่ามากกว่าค่า Chi-Square จากตาราง ที่ probability 0.05 แสดงว่าโอกาสที่จะได้ค่าสังเกตดังกล่าวเกิดขึ้นน้อยกว่า 5 ใน 100 หรือ 1 ใน 20 ครั้ง ซึ่งมีโอกาสน้อยมาก เราก็ไม่ยอมรับว่ามันเป็นไปตามสมมติฐานที่ตั้งไว้ แต่ถ้าค่า Chi-Square ที่ได้มีค่าน้อยกว่า ค่า Chi-Square จากตาราง ที่ probability 0.05 เราก็พอจะยอมรับได้ว่าความแตกต่างที่เกิดขึ้นเป็นเพราะโอกาสที่เป็นไปได้ (chance)

ค่า Chi-Square จะต้องหามาจากจำนวนค่าสังเกตที่ได้มาจากการทดลองโดยตรง ไม่ใช่ค่า percent หรือ ratio มาคำนวณ เช่นเราทดลองแล้วได้ค่าสังเกตออกมาเป็น 8 และ 12 ถ้าคิดเป็น percent จะได้ 40% (หรือ 0.40) และ 60% (หรือ 0.60) ส่วนค่าที่คาดหวัง (expected value) จะได้ 50% และ 50% เราจะใช้ค่า 8 และ 12 (expected value 10 และ 10) มาคำนวณค่า Chi-Square ไม่ใช่ค่า 40% , 60% และ expected value 50% , 50% มาคำนวณ

ตัวอย่างการทดสอบอัตราส่วนทางพันธุศาสตร์ โดยใช้ Chi-Square Test

ตัวอย่างที่ 1

ในการทดลองผสม backcross ระหว่างถั่วลันเตาดันสูงที่เป็น heterozygous กับถั่วลันเตาดันเตี้ย ลูกที่ได้จากการผสม เป็นต้นสูง 260 ต้น เป็นต้นเตี้ย 240 ต้น รวม 500 ต้น

ในทางทฤษฎี เราคาดว่าจะได้ต้นสูง : ต้นเตี้ย ในอัตรา 1:1

Phenotype	Observed value	Expected value	Deviation D = O-E	D ²	D ² /E
ต้นสูง	260	(1/2) × 500 = 250	260-250 = 10	10 ² = 100	100/250 = 0.4
ต้นเตี้ย	240	(1/2) × 500 = 250	240-250 = -10	(-10) ² = 100	100/250 = 0.4
รวม	500	500	0		0.8

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} = 0.8$$

degree of freedom = 2-1 = 1

Probability : 0.30 < P < 0.50 (non significant)

สรุป อัตราส่วนที่ได้จากการทดลองสอดคล้องกับอัตราส่วนทางทฤษฎี ความเบี่ยงเบนที่เกิดขึ้นนั้นเนื่องมาจากโอกาสที่จะเป็นไปได้ที่ลูกที่ได้จากการผสมบางต้นจะเกิดเป็นพวกต้นสูง แทนที่จะเป็นพวกต้นเตี้ย

ตัวอย่างที่ 2

ในถั่วลันเตา ลักษณะใบเลี้ยงสีเหลืองเข้มลักษณะใบเลี้ยงสีเขียว ลักษณะฝักเรียบขม ลักษณะฝักย่น จากการศึกษาลูกชั่วที่สอง จำนวน 994 ต้น มีการกระจายตัวดังนี้ คือ ใบเลี้ยงสีเหลือง ฝักเรียบ 556 ต้น ใบเลี้ยงสีเหลือง ฝักย่น 184 ต้น ใบเลี้ยงสีเขียว ฝักเรียบ 193 ต้น ใบเลี้ยงสีเขียว ฝักย่น 61 ต้น

ให้ใช้ Chi-square ทดสอบว่าลูกชั่วที่ 2 จะมีการกระจายตัวสอดคล้องกับอัตราส่วน 9:3:3:1 หรือไม่

Phenotype	Observed value	Expected value
ใบเลี้ยงสีเหลือง ฝักเรียบ	556	(9/16) × 994 = 559.125
ใบเลี้ยงสีเหลือง ฝักย่น	184	(3/16) × 994 = 186.375
ใบเลี้ยงสีเขียว ฝักเรียบ	193	(3/16) × 994 = 186.375
ใบเลี้ยงสีเขียว ฝักย่น	61	(1/16) × 994 = 62.125
รวม	994	994.0

Phenotype	Deviation D = O-E	D ²
ใบเลี้ยงสีเหลือง ฝักเรียบ	556-559.125 = -3.125	9.765625
ใบเลี้ยงสีเหลือง ฝักย่น	184-186.375 = -2.375	5.640625
ใบเลี้ยงสีเขียว ฝักเรียบ	193-186.375 = 6.625	43.890625
ใบเลี้ยงสีเขียว ฝักย่น	61-62.125 = -1.125	1.265625
รวม	0.0	

Phenotype	D ² /E
ใบเลี้ยงสีเหลือง ฝักเรียบ	9.765625/559.125 = 0.017466
ใบเลี้ยงสีเหลือง ฝักย่น	5.640625/186.375 = 0.030265
ใบเลี้ยงสีเขียว ฝักเรียบ	43.890625/186.375 = 0.235496
ใบเลี้ยงสีเขียว ฝักย่น	1.265625/62.125 = 0.020372
รวม	0.303599

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} = 0.303599$$

degree of freedom = 4-1 = 3

Probability : P > 0.95 (non significant)

สรุป การกระจายตัวของลูกข้าวที่สองเป็นไปตามอัตราส่วน 9:3:3:1 gene ที่ควบคุมลักษณะสีของใบเลี้ยง และ gene ที่ควบคุมลักษณะฝัก เป็นอิสระต่อกัน การถ่ายทอดลักษณะเป็นไปตามกฎการเข้าชุดอย่างอิสระของ gene (law of independent assortment)

Chi-Square Distribution

Degree of freedom	Probability											
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001	
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83	
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82	
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27	
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47	
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52	
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46	
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32	
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12	
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88	
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59	
	Nonsignificant								Significant			

Source : R.A. Fisher and F. Yates, Statistical Tables for Biological, Agricultural and Medical Research (6th edition), Table IV, Oliver & Boyd, Ltd., Edinburgh, by permission of the authors and publishers.

*