10. การแจกแจงแบบไคสแควร์ และการวิเคราะห์ความแปรปรวน

1. บทนำ
2. การใช้ไคสแควร์ทดสอบความเป็นอิสระ
3. การใช้ไคสแควร์ทดสอบสารรูปสนิทดี (χ^2 - test for goodness of fit)
4. การเปรียบเทียบระหว่าง k สัดส่วน (ของการแจกแจงแบบทวินาม)
5. การวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว
6. การวิเคราะห์ความแปรปรวนแบบจำแนก 2 ทาง
7. การอนุมานความแปรปรวนของ 1 ประชากร
8. การอนุมานความแปรปรวนของ 2 ประชากร
9. แบบฝึกหัด

1. บทนำ

ในบทที่ 9 ได้กล่าวถึงวิธีการทดสอบโดยใช้ข้อมูลจากตัวอย่างกลุ่มเดียว หรือ 2 กลุ่มตัวอย่าง เราใช้ตัวอย่าง 1 กลุ่ม เพื่อทดสอบว่า ค่าเฉลี่ยหรือสัดส่วนที่ได้จากตัวอย่างนั้นแตกต่างกับค่าพารามิเตอร์ที่ตั้งข้อสมมุติไว้หรือไม่ ส่วนข้อมูลที่ได้จากตัวอย่าง 2 กลุ่ม เราก็ใช้ทดสอบความแตกต่างของค่าเฉลี่ยหรือสัดส่วนระหว่าง 2 ประชากร

แต่ถ้าเรามีค่าสัดส่วนจากตัวอย่าง 5 กลุ่ม และเราต้องการเปรียบเทียบสัดส่วนเหล่านี้ เราจะใช้วิธีการในบทที่ 9 ไม่ได้ ในบทนี้ จะกล่าวถึงการทดสอบแบบไคสแควร์ซึ่งใช้สำหรับทดสอบว่าประชากรตั้งแต่ 2 กลุ่มขึ้นไป ว่ามีค่าพารามิเตอร์ต่างกันหรือไม่

นอกจากนั้น เรายังใช้การทดสอบแบบไคสแควร์สำหรับทดสอบความเป็นอิสระกันของ 2 คุณลักษณะ

และถ้าเรามีค่าเฉลี่ยจากตัวอย่างมากกว่า 2 กลุ่มขึ้นไป และเราต้องการเปรียบเทียบว่ามาจากประชากรที่มีพารามิเตอร์ต่างกันหรือไม่ เราจะใช้วิธีการทดสอบในบทที่ 9 ไม่ได้ ในบทนี้จะกล่าวถึง การวิเคราะห์ความแปรปรวน ซึ่งเป็นวิธีการทดสอบว่าค่าเฉลี่ยของประชากรตั้งแต่ 2 กลุ่มขึ้นไปมีความแตกต่างกันหรือไม่

นอกจากเราจะสนใจค่าเฉลี่ย และสัดส่วนของประชากรแล้ว ยังมีพารามิเตอร์อีกตัวหนึ่งที่ควรสนใจ คือ ความแปรปรวนของประชากร ในบทนี้ จะได้แสดงการใช้การแจกแจงแบบไคสแควร์สร้างช่วงเชื่อมั่นและทดสอบสมมุติฐานเกี่ยวกับความแปรปรวนของ 1 ประชากร และท้ายสุดจะแสดงการทดสอบสมมุติฐานเกี่ยวกับความแปรปรวนของ 2 ประชากร โดยใช้การแจกแจงแบบ F

แบบฝึกหัด

10.1 เหตุใดจึงต้องใช้การทดสอบแบบไคสแควร์?

10.2 เหตุใดจึงต้องทำการวิเคราะห์ความแปรปรวน?

10.3 เราควรใช้การทดสอบแบบใดสำหรับสภาวการณ์ต่อไปนี้

ก) ต้องการทราบว่า วิธีการส่งเสริมการขาย 3 วิธี จะทำให้จำนวนขายถัวเฉลี่ยแตก ต่างอย่างมีนัยสำคัญ หรือไม่

ข) ต้องการทราบว่ายาสีฟัน X ได้รับความนิยมเท่ากันระหว่างหญิงและชายหรือไม่

ค) ต้องการเปรียบเทียบสัดส่วนลูกค้าที่นิยมผงซักฟอก 4 ชนิด

10.4 จะใช้การแจกแจงแบบใด หรือการทดสอบแบบใดที่เหมาะสมกับการเปรียบเทียบระหว่างกลุ่มต่าง ๆ ต่อไปนี้

ก) เปอร์เซ็นต์แรงงานในกลุ่มอายุ : 16-23, 24-31, 32-39, 40-47, 48-55 และ 56 ขึ้นไป

ข) รายได้เฉลี่ยของกลุ่มอายุ : 16-23, 24-31, 32-39, 40-47, 48-55 และ 56 ขึ้นไป

ค) รายได้เฉลี่ยของหญิงและชายที่มีอายุ 16-56 ปี

ง) ความแปรปรวนหรือการกระจายของรายได้หญิง และชายที่มีอายุ 16-56 ปี

---

2. การใช้ไคสแควร์ทดสอบ

ความเป็นอิสระของ 2 คุณลักษณะ

ตารางคอนทินเจนซี (contingency table)

ตารางคอนทินเจนซี คือ ตาราง 2 ทาง คือ ทางด้านแถว (แนวนอน) และคอลัมน์ (แนวตั้ง) จะต้องมีมากกว่า 1 แถว และมากกว่า 1 คอลัมน์ ดังนั้น ตารางคอนทินเจนซีที่เล็กที่สุดคือตารางที่มี 2 แถว และ 2 คอลัมน์ เรียกว่าขนาด 2×2 หรือ $r \times c$ ในเมื่อ r คือจำนวนแถว และ c คือจำนวนคอลัมน์ เราใช้ตารางคอนทินเจนซีสำหรับข้อมูลที่จำแนกตามคุณลักษณะ คือ ให้คุณลักษณะ A อยู่ทางด้านแถว และคุณลักษณะ B อยู่ด้านคอลัมน์ และจุดประสงค์ขั้นต่อไปคือการตรวจสอบว่าคุณลักษณะ A และ B มีความสัมพันธ์กัน หรือเป็นอิสระกัน เช่น ต้องการตรวจสอบความสัมพันธ์ระหว่างการสูบบุหรี่กับการเป็นโรคปอด จะเก็บข้อมูลจำแนกในตารางขนาด 2×2 ดังนี้

		B เป็นโรคปอด	 ไม่เป็นโรคปอด	
A	สูบบุหรี่	n_{11}	n_{12}	$n_{1.}$
	ไม่สูบบุหรี่	n_{21}	n_{22}	$n_{2.}$
		$n_{.1}$	$n_{.2}$	n

ตัวอย่าง ต้องการทดสอบว่า ผลการทำงานของพนักงานมีส่วนเกี่ยวข้องกับระดับการศึกษาหรือไม่ จากพนักงานที่สุ่มมา 100 คน เมื่อจำแนกตามระดับการศึกษา มี 40 คน ที่ผ่านการศึกษาระดับวิทยาลัย อีก 60 คนไม่ผ่าน และเมื่อจำแนกตามผลงาน พบว่า ในบรรดาผู้จบวิทยาลัย มี 15

คน ที่ให้ผลการทำงานดี ส่วนที่เหลืออีก 25 คน ให้ผลงานไม่ดีนัก ส่วนผู้ที่ไม่ผ่านวิทยาลัย มี 15 คนที่มีผลงานในขั้นดี ที่เหลือ 45 คน ให้ผลงานไม่ดีนัก ดังนั้น จึงมีผู้ให้ผลงานในขั้นดีทั้งหมด 30 คน และ 70 คนให้ผลงานไม่ดีนัก จึงรวมข้อมูลนี้ใส่ในตารางคอนทินเจนซีขนาด 2 × 2 ดังนี้

B = ความรู้

	การทำงาน	จบวิทยาลัย	ไม่จบวิทยาลัย	รวม
A	ดี	15	15	30
	ไม่ดีนัก	25	45	70
	รวม	40	60	100

ก่อนที่จะแสดงการทดสอบของตัวอย่างข้างต้น จะแสดงขั้นตอนการทดสอบโดยทั่วไปก่อน ดังนี้

1. Ho : คุณลักษณะ A และ B เป็นอิสระกัน
2. Ha : คุณลักษณะ A และ B ไม่เป็นอิสระกัน
3. กำหนดระดับนัยสำคัญ α
4. เขตวิกฤตจะอยู่ภายใต้โค้งการแจกแจงแบบไคสแควร์ด้านขวามือที่มี df = (r - 1) (c - 1) นั่นคือจะปฏิเสธเมื่อ $X^2 > \chi^2_{(r-1)(c-1),\ \alpha}$
5. ตัวสถิติที่ใช้ทดสอบ คือ

$$X^2 = \sum_i^r \sum_j^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad \begin{array}{l} i = 1, 2, \ldots, r \\ j = 1, 2, \ldots, c \end{array}$$

$$\boxed{E_{ij} = \frac{R_i \times C_j}{n}}$$

ในเมื่อ

O_{ij} = observed frequency ของลักษณะในแถวที่ i และคอลัมน์ที่ j

E_{ij} = expected frequency ของลักษณะในแถวที่ i และคอลัมน์ที่ j

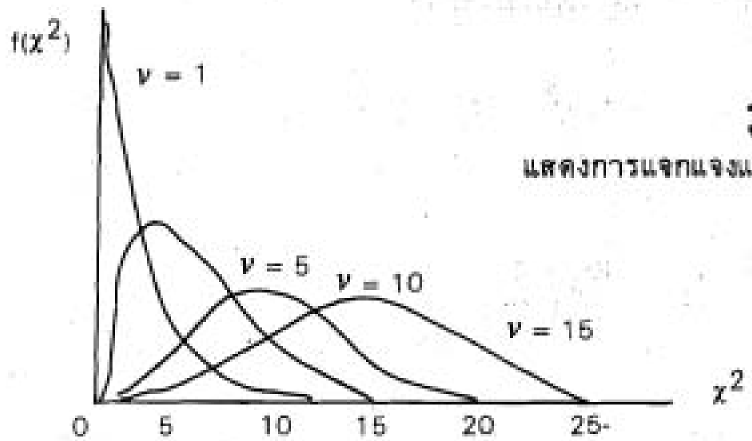
= Ri X Cj/n

6. ปฏิเสธ Ho ถ้า X^2 อยู่ในเขตวิกฤต

<u>ตารางการแจกแจงแบบ χ^2</u>

การเปิดตารางการแจกแจงของไคสแควร์ (χ^2) จะเหมือนกับการเปิดตาราง t คือต้องทราบค่า df(v) และค่า α รูปร่างของ χ^2 จะเบ้ไปทางด้านขวามือ และมีรูปร่างเปลี่ยนตามค่า v และ

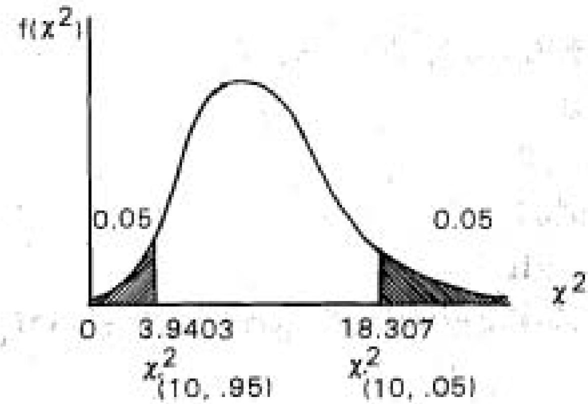
เมื่อ ν มีค่าโตมากการแจกแจงแบบ χ^2 จะสามารถประมาณได้โดยโค้งปกติ การแจกแจงแบบ χ^2 เป็นการแจกแจงแบบต่อเนื่อง และมีฐานนิยมเพียงแห่งเดียว



รูปที่ 10.1
แสดงการแจกแจงแบบ χ^2 ที่มี df 1, 5, 10, 15

ตารางที่ 10.1
แสดงส่วนหนึ่งของตาราง χ^2

ν \ α	พื้นที่ด้านปลายขวามือ						
	0.99	...	0.95	....	0.05	....	0.01
⋮							
10	2.55821		3.9403		18.3070		23.2093
⋮							



รูปที่ 10.2
แสดงการแจกแจงแบบ χ^2 ที่มี df = 10

ดังนั้น จากตัวอย่าง เราจะตั้งสมมุติฐานได้ดังนี้

1) Ho : ผลการทำงานและระดับการศึกษาเป็นอิสระกัน

2) Ha : ผลการทำงาน และระดับการศึกษาไม่เป็นอิสระกัน

3) $\alpha = 0.05$

4) $df = (r-1)(c-1) = (2-1)(2-1) = 1$, $\chi^2_{1,\ .05} = 3.84$

นั่นคือจะปฏิเสธ Ho เมื่อ

$$\chi^2 > 3.84$$

5. ตัวสถิติที่ใช้ทดสอบคือ $\chi^2 = \sum_i^r \sum_j^c \dfrac{(O_{ij} - E_{ij})^2}{E_{ij}}$

Oij คือ ข้อมูลในตารางที่เก็บมา ดังนี้

O_{11}	O_{12}
O_{21}	O_{22}

	b_1 จบวิทยาลัย	b_2 ไม่จบวิทยาลัย	
a_1 = ผลงานดี	15	15	30
a_2 = ผลงานไม่ดี	25	45	70
	40	60	100

ส่วนค่า Eij จะต้องหามาโดยสมมุติว่าลักษณะทั้ง 2 เป็นอิสระกันตามที่กำหนดไว้ในสมมุติฐานว่างเปล่า ซึ่งเราทราบตามกฎความน่าจะเป็นว่า

ถ้า A และ B เป็นอิสระกัน

$P(A \cap B) = P(A).P(B)$

ในขณะนี้เรามีเหตุการณ์ A และ B ที่เกิดร่วมกัน 4 เหตุการณ์ คือ

a_1b_1 = เป็นผู้จบวิทยาลัยและผลงานดี

a_1b_2 = เป็นผู้ไม่จบวิทยาลัยและผลงานดี

a_2b_1 = เป็นผู้จบวิทยาลัยและผลงานไม่ดี

a_2b_2 = เป็นผู้ไม่จบวิทยาลัยและผลงานไม่ดี

ความน่าจะเป็นที่จะเกิดเหตุการณ์ทั้ง 4 คู่ลำดับนั้นจะต้องสอดคล้องกับกฎความเป็นอิสระเชิงสถิติ นั่นคือ

$P(a_1b_1) = P(a_1).P(b_1) = (\frac{30}{100})(\frac{40}{100})$

$P(a_1b_2) = P(a_1).P(b_2) = (\frac{30}{100})(\frac{60}{100})$

$P(a_2b_1) = P(a_2).P(b_1) = (\frac{70}{100})(\frac{40}{100})$

$P(a_2b_2) = P(a_2).P(b_2) = (\frac{70}{100})(\frac{60}{100})$

ดังนั้น เมื่อเราสามารถหาความน่าจะเป็นของเหตุการณ์ทั้งหมดภายใต้ข้อสมมุติว่าลักษณะทั้ง 2 เป็นอิสระ กันได้แล้ว เราจึงหาจำนวนคาดหมายได้โดยนำความน่าจะเป็นของแต่ละเหตุการณ์คูณจำนวนคนทั้งหมด (n) นั่นคือ

$$E_{11} = P(a_1b_1)\times n = (\frac{30}{100})(\frac{40}{100})100$$
$$= \frac{30 \times 40}{100} = \frac{R_1 \times C_1}{n} = 12 \text{ คน}$$

$$E_{12} = P(a_1b_2)\times n = (\frac{30}{100})(\frac{60}{100})100$$
$$= \frac{30 \times 60}{100} = \frac{R_1 \times C_2}{n} = 18 \text{ คน}$$

$$E_{21} \quad P(a_2b_1)\times n = (\frac{70}{100})(\frac{40}{100})100$$
$$= \frac{70 \times 40}{100} = \frac{R_2 \times C_1}{n} = 28 \text{ คน}$$

$$E_{22} = P(a_2b_2)\times n = (\frac{70}{100})(\frac{60}{100})100$$
$$= \frac{70 \times 60}{100} = \frac{R_2 \times C_2}{n} = 42 \text{ คน}$$

ถ้าให้ R_i คือผลรวมของแถวที่ i

C_j คือผลรวมของคอลัมน์ที่ j

เราจะได้กฎการหาจำนวนคาดหมายของแถวที่ i และคอลัมน์ j ดังนี้

$$\boxed{E_{ij} = \frac{R_i \times C_j}{n}} \quad , i = 1, 2, ..., r \qquad j = 1, 2, ..., c$$

ค่า E_{ij} ทั้งหลาย ควรจัดรวบรวมใส่ตารางคอนทินเจนซี เพื่อสะดวกในการคำนวณค่าสถิติ X^2

	จบ วิทยาลัย	ไม่จบ วิทยาลัย	
ผลงาน ดี	15	15	30
ผลงาน ไม่ดี	25	45	70
	40	60	100

ตารางแสดงค่า O_{ij}

	จบ วิทยาลัย	ไม่จบ วิทยาลัย	
ผลงาน ดี	12	18	30
ผลงาน ไม่ดี	28	42	70
	40	60	100

ตารางแสดงค่า E_{ij}

พึงสังเกตความคล้ายคลึงกันของ 2 ตารางนี้ จะเห็นว่าค่าที่มุมตารางที่เรียกว่า marginal total เท่ากัน ค่าที่ต่างกันคือค่าในตาราง สำหรับตาราง E_{ij} ได้ทำการจัดสรรความถี่โดยสมมุติว่าลักษณะทั้ง 2 เป็นอิสระกัน อีกข้อที่ควรสังเกตคือการที่มี df = 1 แสดงว่ามีเพียงตัวเดียวใน 4 ตัว นั้นที่เป็นอิสระ เพราะเมื่อเรากำหนดตัวใดตัวหนึ่งเพียง 1 ตัว ที่เหลืออีก 3 ตัว เราจะหาได้ทันทีโดยนำไปหักออกจากค่า marginal total ซึ่งเราทราบล่วงหน้าจากตาราง O_{ij} แล้ว เช่น ถ้าใส่ $E_{11} = 12$ จะหา $E_{12} = 30 - 12 = 18$, $E_{21} = 40 - 12 = 28$ และ $E_{22} = 60 - 18 = 42$ หรือ $70 - 28 = 42$

เมื่อทราบค่า O_{ij} และ E_{ij} ทั้งหมดแล้ว จึงคำนวณค่าสถิติ X^2 ดังนี้

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

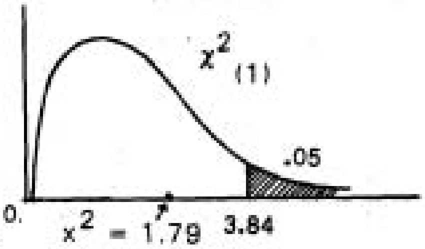
$$= \frac{(15-12)^2}{12} + \frac{(15-18)^2}{18} + \frac{(25-28)^2}{28} + \frac{(45-42)^2}{42}$$

$$= \frac{(3)^2}{12} \; \frac{(-3)^2}{18} + \frac{(-3)^2}{28} + \frac{(3)^2}{42}$$

$$= .75 + .50 + .32 + .22 = 1.79$$

โปรดสังเกตตัวตั้ง คือผลรวมของส่วนเบี่ยงเบน ก่อนยกกำลังสอง หรือ $\Sigma\Sigma(O_{ij} - E_{ij})$ จะเห็นว่าได้ผลรวมเป็น 0 ตามหลักผลรวมของค่าเบี่ยงเบนจากค่าเฉลี่ยต้องเป็นศูนย์ จึงเป็นแนวทางให้เราตรวจสอบความถูกต้องได้ทางหนึ่ง

เนื่องจากค่า x^2 = 1.79 ไม่อยู่ในเขตวิกฤต 5% นั้น เราจึงยังปฏิเสธ Ho ไม่ได้ นั่นคือยังไม่มีหลักฐานเพียงพอที่จะสรุปว่า ผลการทำงานและระดับการศึกษาของพนักงานมีความเกี่ยวข้องกัน



(p -value > .05 จึงไม่ปฏิเสธ Ho)

แบบฝึกหัด

10.5 เราจะทราบ df สำหรับเปิดตาราง x^2 เพื่อทดสอบความเป็นอิสระได้อย่างไร?

10.6 ในการทดสอบความเป็นอิสระ ถ้าค่า x^2 ที่คำนวณได้เล็กมาก เราจะปฏิเสธสมมุติว่างเปล่าได้ไหม? เพราะเหตุใด

10.7 ในการทดสอบความเป็นอิสระ เราจะปฏิเสธ Ho ได้ไหมถ้าค่า x^2 ที่คำนวณได้เป็นค่าที่โตมาก? เพราะเหตุใด?

10.8 ตารางคอนทินเจนซีคืออะไร สมมุติฐานของการทดสอบว่าอย่างไร?

10.9 ในการลงคะแนนเสียงกฎหมายฉบับหนึ่ง มีผู้แทนที่ออกเสียงสนับสนุน และคัดค้านโดยจำแนกตามพรรคการเมือง ดังนี้

การออกเสียง	พรรคร่วมรัฐบาล	พรรคฝ่ายค้าน
คัดค้าน	250	200
สนับสนุน	400	150

จงทดสอบด้วยระดับนัยสำคัญ .05 ว่า ผลการออกเสียงไม่มีความสัมพันธ์กับพรรคการเมือง (x^2 = 32.078, $x^2_{(1),\ .05}$ = 3.84, ปฏิเสธ Ho)

10.10 ต้องการทดสอบว่า ความสัมฤทธิ์ผลในการทำงานเป็นอิสระกับความสัมฤทธิ์ผลทางการศึกษา ได้สุ่มพนักงานมา 100 คน และจำแนกใส่ตาราง 3 x 3 คอนทินเจนซี ดังนี้

ผลการทำงาน	ความสัมฤทธิ์ผลในการศึกษา			รวม
	A	B	C หรือต่ำกว่า	
ดีมาก	10	5	5	20
ปานกลาง	20	12	8	40
เลว	20	13	7	40
รวม	50	30	20	100

ถ้าใช้ α = .05 จะสรุปว่าความสัมฤทธิ์ผลในการงานเป็นอิสระกับความสัมฤทธิ์ผลด้านการศึกษาได้ไหม? (X^2 = 0.63, ไม่ ปฏิเสธ Ho)

10.11 ผู้ผลิตสินค้าต้องการทราบว่า ความนิยมสินค้าแบบต่างมีความสัมพันธ์กับเพศของลูกค้าหรือไม่ จากการสุ่มผู้ซื้อสินค้าของบริษัทมา 1,000 ราย ได้ข้อมูลดังนี้

เพศ	แบบที่นิยมซื้อ			รวม
	1	11	111	
ชาย	100	100	200	400
หญิง	300	150	150	600
รวม	400	250	350	1,000

จงใช้ α = 0.05 ทดสอบว่า เพศ และความนิยมไม่เกี่ยวข้องกัน (X^2 = 80.35, ปฏิเสธ Ho)

10.12 ผู้ผลิตวัคซีนป้องกันโรคหวัด ต้องการทดสอบวัคซีนชนิดใหม่โดยแบ่งผู้ทดลองเป็น 2 กลุ่ม และฉีดยาให้กลุ่มหนึ่งซึ่งมี 30 คน ส่วนอีก 20 คนไม่ฉีดยาโดยถือเป็นกลุ่ม "ควบคุม" (control) เพื่อใช้เปรียบเทียบได้ข้อมูลดังนี้

ผล	ฉีดวัคซีน	ไม่ฉีดวัคซีน	รวม
เป็นหวัด	10	10	20
ไม่เป็นหวัด	20	10	30
รวม	30	20	50

วัคซีนมีผลในการรักษาโรคหวัดไหม? $\alpha = 0.05$ ($X^2 = 1.39$, ไม่ปฏิเสธ Ho)

3. การใช้ไคสแควร์ทดสอบสารรูปสนิทดี

Chi-square as a test of goodness of fit

: testing the appropriateness of a distribution.

เราสามารถใช้การทดสอบแบบไคสแควร์ตรวจดูว่าข้อมูลมาจากประชากรแบบใด เช่น ปัวซอง ทวินาม พหุนาม หรือ แบบโค้งปกติ โดยเราใช้ตัวสถิติเดิมทดสอบ คือ $X^2 = \Sigma (O_i - E_i)^2/E_i$ ในเมื่อ O_i คือข้อมูลที่เราเก็บมาจากตัวอย่าง และต้องอยู่ในรูปจำนวนที่นับได้ (ไม่ใช่สัดส่วน) ส่วน E_i คือ จำนวนคาดหมายจากสมมุติฐานว่างเปล่า กล่าวคือ เราจะเลือกการแจกแจงที่เหมาะสมก่อน และตั้งสมมุติฐานว่างเปล่าว่าข้อมูลมาจากประชากรที่มีการแจกแจงตามที่เราเลือก และเราจะคำนวณความน่าจะเป็นของเหตุการณ์ต่าง ๆ เพื่อหาค่า E_i ถ้า O_i และ E_i มีความขัดแย้งกันน้อย ค่าสถิติ x^2 จะมีค่าน้อยด้วย จะยอมรับ Ho และสรุปว่า ข้อมูลมีการแจกแจงตามที่เราอ้างไว้ใน Ho ในทางตรงข้าม ถ้าค่า O_i และ E_i มีความขัดแย้งกันสูง ค่าสถิติ x^2 จะโตมากจนตกอยู่ในเขตวิกฤต ดังนั้น เราจะปฏิเสธ Ho และสรุปว่าข้อมูลไม่ได้มีการแจกแจงตามที่เราอ้างไว้ใน Ho

ตัวอย่าง 1 ในการคัดเลือกพนักงานของบริษัทแห่งหนึ่ง ใช้วิธีให้กรรมการ 3 คน สัมภาษณ์ และให้คะแนนว่าผ่านหรือไม่ผ่าน ผู้สมัครทุกคนต้องให้กรรมการทั้ง 3 คนสัมภาษณ์ ฝ่ายบุคลากรคาดว่า จำนวนผู้ที่ผ่านการสอบสัมภาษณ์น่าจะมีการแจกแจงแบบทวินามที่มีพารามิเตอร์ $\pi = .4$ (คาดว่ามี 40% ของผู้สมัครที่สอบผ่าน) จงทดสอบด้วยระดับนัยสำคัญ .20 ข้อมูลที่เก็บได้จากการสัมภาษณ์ผู้สมัคร 100 คน มีดังนี้

จำนวนกรรมการที่ให้คะแนนสอบผ่าน (x)	จำนวนผู้สมัคร ($f_i = O_i$)
0	18
1	47
2	24
3	11
	100

1. Ho : ข้อมูลมีการแจกแจงแบบทวินามที่มี π = .40 หรือ การแจกแจงแบบทวินามที่มี π = .40 ปรับเข้า (fit) กับข้อมูลได้อย่างดี

2. Ha : ข้อมูลมีการแจกแจงแบบทวินามที่มี π = .40 หรือ การแจกแจงแบบทวินามที่มี π = .40 ปรับเข้า (fit) กับข้อมูลไม่ดีนัก

3. $\alpha = .20$

4. เขตวิกฤตอยู่ด้านขวามือของโค้ง χ^2 ที่มี $df = k - 1 = 4 - 1 = 3$, k คือจำนวนเหตุการณ์ที่เป็นไปได้ จากตาราง $\chi^2_{3,\ .20} = 4.642$ ดังนั้น เขตปฏิเสธ Ho คือ เมื่อ $\chi^2 > 4.642$

5. คำนวณค่าสถิติ $\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$

O_i คือข้อมูลที่เก็บมา คือ 18, 47, 24, 11

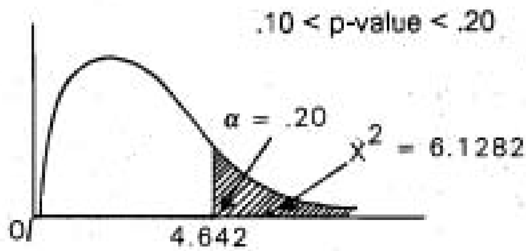
E_i คือจำนวนคาดหมาย เมื่อ x มีค่าเป็น 0, 1, 2, 3 จึงต้องหาความน่าจะเป็น คือหา $P(X = 0)$, $P(X = 1)$, $P(X = 2)$ และ $P(X = 3)$ ซึ่งจะยึดถือการแจกแจงที่อ้างใน Ho คือการแจกแจงแบบทวินามที่มี $n = 3$ และ π = .40 จะได้ความน่าจะเป็นดังนี้

$P(X = 0) = .2160$	จำนวนคาดหมาย	คือ $21.6 \times 100 = 21.6 =$	22 คน
$P(X = 1) = .4320$	จำนวนคาดหมาย	คือ $.4320 \times 100 = 43.2 =$	43 คน
$P(X = 2) = .2880$	จำนวนคาดหมาย	คือ $.2880 \times 100 = 28.8 =$	29 คน
$P(X = 3) = .0640$	จำนวนคาดหมาย	คือ $.064 \times 100 = 6.4 =$	6 คน
1.0000			100 คน

ควรนำค่า O_i และ E_i จัดใส่ตารางเพื่อคำนวณค่าสถิติ X^2 ดังนี้

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
18	22	-4	16	.7273
47	43	4	16	.3721
24	29	-5	25	.8621
11	6	5	25	4.1667
100	100	0	$X^2 = \Sigma(O_i - E_i)^2/E_i \longrightarrow$	6.1282

6. ค่า $X^2 = 6.1282$ อยู่ในเขตวิกฤต จึงปฏิเสธ Ho และสรุปว่า การแจกแจงแบบทวินามที่มี $\pi = .40$ ไม่สอดคล้อง (ไม่ fit) กับข้อมูลที่เก็บมา



ตัวอย่าง 2 เพื่อจะทดสอบว่าเป็นลูกเต๋าที่สมดุลย์ จึงได้ทดลองโยนลูกเต๋านั้น 300 ครั้ง ได้ผลการทดลองดังนี้

หน้าที่ขึ้น	1	2	3	4	5	6
จำนวนครั้ง	35	40	32	60	68	65

จะสรุปว่าเป็นลูกเต๋าสมดุลย์ด้วยระดับนัยสำคัญ $\alpha = .01$ ได้หรือไม่?

1) $Ho : \pi_i = \frac{1}{6}, i = 1, 2, ..., 6$ หมายความว่า ข้อมูลมีการแจกแจงแบบยูนิฟอร์ม เพราะโอกาสที่จะเกิดเหตุการณ์ทั้ง 6 อันเท่ากัน ในขณะเดียวกันข้อมูลมีการแจกแจงแบบพหุนามที่มีพารามิเตอร์ $\pi_i = \frac{1}{6}$ ด้วย ถ้ายอมรับ Ho จะสรุปว่าเป็นลูกเต๋าที่สมดุลย์

2) Ha : $\pi_i \neq \frac{1}{6}$

หมายความว่า ข้อมูลไม่ได้มีการแจกแจงแบบยูนิฟอร์ม เพราะโอกาสการเกิดเหตุการณ์ ทั้ง 6 ไม่เท่ากัน ในขณะเดียวกันเป็นการทดสอบว่าข้อมูลมีการแจกแจงแบบพหุนามซึ่งมีพารามิเตอร์ π_i ไม่เท่ากันทั้งหมด แต่ไม่ระบุว่าแต่ละค่ามีค่าเป็นเท่าใด ระบุเพียงว่า ไม่ใช่ $\frac{1}{6}$ สำหรับทุกค่าเท่านั้น ถ้ายอมรับ Ha จะสรุปว่าเป็นลูกเต๋าที่ไม่สมดุลย์

3. $\alpha = .01$

4. จะปฏิเสธ Ho เมื่อ $X^2 > \chi^2_{(k-1),\alpha}$ คือ $X^2 > \chi^2_{(6-1),.01} = 15.0863$

5. ตัวสถิติที่ทดสอบ คือ

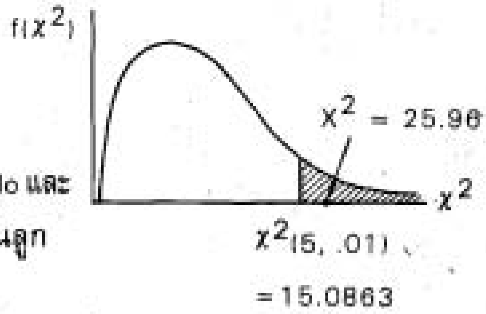
$$X^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i}$$

ค่า O_i คือ จำนวนครั้งของหน้าต่าง ๆ ที่ได้จากการทดลอง 300 ครั้ง
ค่า E_i จะต้องหาจากข้อสมมุติใน Ho ซึ่งอ้างว่า $\pi_i = \frac{1}{6}$ ดังนั้น จำนวนคาดหมายของเหตุการณ์ $i = n\pi_i = 300(\frac{1}{6}) = 50$

เราจะแสดงการคำนวณค่าสถิติ X^2 ได้ดังนี้

หน้า	จำนวนครั้ง O_i	ความน่าจะเป็น π_i	จำนวนคาดหมาย $E_i = n\pi_i$	O- E	$(O-E)^2$	$(O-E)^2/E$
1	35	1/6	50	-15	225	4.50
2	40	1/6	50	-10	100	2.00
3	32	1/6	50	-18	324	6.48
4	60	1/6	50	10	100	2.00
5	68	1/6	50	18	324	6.48
6	65	1/6	50	15	225	4.50
รวม	300	1.00	300	0		25.96

p-value < .01 จึงปฏิเสธ Ho



6. ค่าสถิติ $\chi^2 = 25.96$ อยู่ในเขตวิกฤตจึงปฏิเสธ Ho และสรุปว่าข้อมูลไม่มีการแจกแจงแบบยูนิฟอร์ม คือเป็นลูกเต๋าที่ไม่สมดุลย์

ตัวอย่าง 3 ถ้าเราจำแนกผู้มีสิทธิออกเสียง ตามระดับความรู้เป็น 5 กลุ่ม แต่ละกลุ่มไม่มีผลร่วมกัน (mutually exclusive) และจากสถิติเมื่อ 20 ปีก่อนเชื่อว่า ผู้มีการศึกษาจบชั้น ป.7 หรือต่ำกว่ามี 35%, มศ.1-มศ. 3 มี 30% มศ. 4 -มศ. 5 มี 20% ได้เข้าศึกษาในมหาวิทยาลัย 10% และจบมหาวิทยาลัย 5% เราต้องการทดสอบว่าผู้มีสิทธิออกเสียงยังคงมีอัตราส่วนจำแนกตามระดับการศึกษาเหมือนกับเมื่อ 20 ปีก่อนหรือไม่ จากการสุ่มตัวอย่างผู้มีสิทธิออกเสียงมา 1,000 ราย มีจำนวนในกลุ่มต่าง ๆ เป็น 315, 270, 230, 125 และ 60 คนตามลำดับ ถ้าใช้ระดับนัยสำคัญ 5% จะสรุปว่าการแจกแจงของผู้มีสิทธิออกเสียงยังเหมือนเดิมไหม?

1) Ho : $\pi_1 : \pi_2 : \pi_3 : \pi_4 : \pi_5 = .35 : .30 : .20 : .10 : .05$

Ho : $\pi_1 = 35$, $\pi_2 = .30$, $\pi_3 = .20$, $\pi_4 = .10$, $\pi_5 = .05$

ในเมื่อ π_i คือสัดส่วนผู้มีสิทธิออกเสียงในกลุ่มการศึกษาต่าง ๆ

ถ้ายอมรับ Ho แสดงว่าอัตราส่วนไม่เปลี่ยนแปลง

2) Ha : $\pi_1 : \pi_2 : \pi_3 : \pi_4 : \pi_5 \neq .35 : .30 : .20 : .10 : .05$

ถ้ายอมรับ Ha แสดงว่าอัตราส่วนมีการเปลี่ยนแปลงจากเดิม

3) $\alpha = .05$

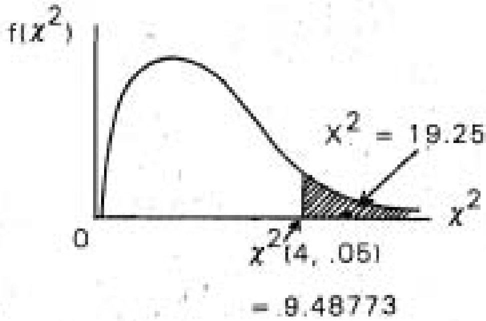
4) จะปฏิเสธ Ho เมื่อ $\chi^2 > \chi^2_{(5-1),\ .05} = 9.48773$

5) คำนวณค่าสถิติ $X^2 = \Sigma_{i=1}^{5}(O_i - E_i)^2/E_i$

การศึกษา	O_i	π_i	$E_i = n\pi_i$	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
ป 7 หรือต่ำกว่า	315	.35	350	-35	1225	3.00
มศ. 1 - มศ. 3	270	.30	300	-30	900	3.50
มศ. 4 - มศ. 5	230	.20	200	30	900	4.50
เข้ามหาวิทยาลัย	125	.10	100	25	625	6.25
จบมหาวิทยาลัย	60	.05	50	10	100	2.00
	1,000	1.00	1,000	0		19.25

6. ค่าสถิติ $X^2 = 19.25$ อยู่ในเขตวิกฤต จึงปฏิเสธ Ho และสรุปว่าอัตราส่วนแตกต่างไปจากเดิมอย่างมีนัยสำคัญ

(p-value < .05 จึงปฏิเสธ Ho)



มีข้อควรระวังเกี่ยวกับการทดสอบแบบไคสแควร์ คือ

1. ตัวเลขที่ใช้คำนวณต้องเป็นจำนวนความถี่ ไม่ใช่สัดส่วนหรือเปอร์เซ็นต์

2. จำนวนคาดหมายของเหตุการณ์แต่ละอันไม่ควรต่ำกว่า 5 ถ้าน้อยกว่า 5 ควรรวมเหตุการณ์นั้นกับชั้นถัดไป เช่น ในตัวอย่างที่ 3 สมมุติว่าสุ่มตัวอย่างมาเพียง 80 คน จำนวนแต่ละกลุ่มการศึกษาคือ 26, 22, 19, 10 และ 3 ตามลำดับ และเมื่อหาค่าคาดหมายแต่ละกลุ่มจะได้ 23, 24, 16, 8 และ 4 จะเห็นว่าค่าคาดหมายของกลุ่มการศึกษาสุดท้ายคือพวกจบมหาวิทยาลัยมีเพียง 4 คน ซึ่งน้อยกว่า 5 จึงต้องยุบรายการนี้ รวมกับรายการถัดไป ดังนี้

		ป.7 หรือต่ำกว่า	มศ.1 - มศ.3	มศ.3-มศ.5	มหาวิทยาลัย	รวม
O_i	=	26	22	19	10 + 3 = 13	80
E_i	=	28	24	16	8 + 4 = 12	80

ค่าสถิติ x^2 ที่คำนวณได้ต้องเทียบกับค่า x^2 ที่มี df = 4 - 1 = 3

3. การหา df ให้ยึดหลักเบื้องต้น = k - 1 คือ จำนวนรายการ - 1 และถ้าต้องประมาณพารามิเตอร์ตัวใด จะต้องลด df เท่ากับจำนวนพารามิเตอร์ด้วย เช่นในตัวอย่างที่ 1 ถ้าไม่ได้อ้างว่าเป็นการแจกแจงแบบทวินามที่มี π = .4 เราต้องหาค่าประมาณของ π ดังนั้น df จะหายไปอีก 1 df จึงเหลือ (4 - 1) - 1 = 2 df

แบบฝึกหัด

10.13 ในการทดสอบการแจกแจงของข้อมูล (test of goodness of fit) เรามีวิธีการหา df อย่างไร?

10.14 ในการทดสอบการแจกแจงของข้อมูล ถ้า x^2 ที่คำนวณได้เป็นค่าที่เล็กเกินไปเราจะปฏิเสธ Ho ไหม? เพราะเหตุใด?

10.15 ในการทดสอบการแจกแจงของข้อมูล ถ้าเราคำนวณค่า x^2 ได้เป็นค่าที่ใหญ่เกินไป เราจะปฏิเสธ Ho ไหม? เพราะเหตุใด?

10.16 ในการทดสอบการแจกแจงของข้อมูล เราจะใช้การทดสอบแบบ 2 ด้าน คือ มีเขตวิกฤตอยู่ 2 ด้านได้ไหม? จงอธิบาย

10.17 ตัวเลขในตารางเลขสุ่ม (random digit table) ควรจะมีลักษณะไม่เอียงเฉเนื่องจากผู้สร้างตารางได้พยายามให้เลขทุกตัวมีโอกาสปรากฏตัวเท่า ๆ กัน เพื่อที่จะทดสอบคุณสมบัติข้อนี้จึงได้ทำการสุ่มมา 100 จำนวน จากตารางเลขสุ่มและนับจำนวนครั้งที่แต่ละตัวปรากฏมีดังนี้

เลข	0	1	2	3	4	5	6	7	8	9	รวม
จำนวนครั้ง	8	11	10	14	7	12	6	9	13	10	100

เราจะปฏิเสธว่าตัวเลขในตารางไม่เป็นแบบสุ่มที่ระดับนัยสำคัญ .05 ไหม?
x^2 = 6.0, ยอมรับ Ho!

10.18 จำนวนอุบัติเหตุในกรุงเทพฯ ในสัปดาห์หนึ่งมีดังนี้

วัน	จำนวนอุบัติเหตุร้ายแรง
อาทิตย์	28
จันทร์	12
อังคาร	10
พุธ	7
พฤหัส	8
ศุกร์	11
เสาร์	24
รวม	100

จงทดสอบว่า อุบัติเหตุในวันสุดสัปดาห์เป็นวันละ 25% ส่วนวันธรรมดาวันละ 10% ด้วย $\alpha = .025$ ($X^2 = 2.2$, ยอมรับ Ho)

10.19 โยนเหรียญสมดุลย์ 4 อัน พร้อม ๆ กัน 160 ครั้ง และนับจำนวนเหรียญที่หงายด้านหัว ได้ผลดังนี้

จำนวนหัว	0	1	2	3	4	รวม
ความถี่	16	35	55	48	6	160

จงใช้ $\alpha = .05$ ทดสอบว่าเป็นเหรียญสมดุลย์ทั้ง 4 เหรียญ ($X^2 = 7.842$ ยอมรับ Ho)

10.20 นักชีววิทยาได้ทำการผสมพันธุ์ถั่วลันเตา ได้ลูกผสมที่มีลักษณะต่าง ๆ ดังนี้

ลำต้นสูง และมีสีสวยงาม	186	ต้น
ลำต้นสูง และไม่มีสี	66	ต้น
ลำต้นเตี้ยแคระและมีสีสวยงาม	54	ต้น
ลำต้นเตี้ยแคระและไม่มีสี	14	ต้น

ถ้าตามหลักพันธุกรรมซึ่งท่านเมนเดลได้สร้างทฤษฎีไว้ ลักษณะต่าง ๆ ดังกล่าวจะต้องเกิดในอัตราส่วน 9 : 3 : 3 : 1 ตามลำดับ ให้ทดสอบว่าผลการทดลองเป็นไปตามทฤษฎีของเมนเดลหรือไม่? $\alpha = .01$ ($X^2 = 3.2$, ผลการทดลองเป็นไปตามทฤษฎีของเมนเดล)

10.21 ในการนำสินค้าตัวใหม่เข้าสู่ตลาด ผู้ผลิตได้แบ่งตลาดผู้ซื้อออกเป็น 10 ส่วน โดยคาดว่าแต่ละส่วนจะมีจำนวนประชากร และผู้มีอำนาจซื้อไม่ต่างกัน และได้ทำการโฆษณาสินค้าโดยสื่อโฆษณาหลายชนิด โดยขอให้ผู้ที่สนใจผลิตภัณฑ์ใหม่ส่งแบบสอบถามมายังบริษัทว่าจะติดต่อกับตัวแทนจำหน่ายของท้องที่ตนได้ที่ใด ในจำนวนแบบสอบถามที่ส่งกลับคืนมายังผู้ผลิต 200 ใบ แยกเป็นของแต่ละท้องที่ได้ดังนี้

ท้องที่	1	2	3	4	5	6	7	8	9	10	รวม
จำนวนใบสอบถาม	22	23	18	16	21	17	19	23	20	21	200

จงทดสอบว่า ทั้ง 10 ท้องที่มีผลตอบสนองเท่ากัน ด้วย α = .025 (X^2 = 2.7, ยอมรับ Ho)

10.22 จำนวนอุบัติเหตุต่อสัปดาห์บนทางหลวงสายหนึ่งซึ่งเก็บสถิติรวม 80 สัปดาห์ มีดังนี้

จำนวนอุบัติเหตุร้ายแรงต่อสัปดาห์	0	1	2	3	4 หรือมากกว่า	รวม
ความถี่	3	22	33	16	6	80

ให้ทดสอบว่าข้อมูลมีการแจกแจงแบบปัวซอง ด้วย α = .05 (ต้องประมาณค่า μ ดังนั้น df จึงหายไปอีก 1 df X^2 = 14.353, ปฏิเสธ Ho)

10.23 โยนเหรียญ 6 อันพร้อมกัน 1,280 ครั้ง ได้ผลดังนี้

จำนวนหัว	0	1	2	3	4	5	6	รวม
ความถี่	26	140	274	420	290	112	18	1,280

ให้ทดสอบว่าเป็นเหรียญสมดุลย์ α = .01 (X^2 = 9.44, ยอมรับ Ho)

10.24 ถ้าใช้ α = .10 จะสรุปว่าข้อมูลต่อไปนี้มีการแจกแจงแบบปัวซองที่มี μ = 2 ไหม?

จำนวนลูกค้าต่อชั่วโมง	0	1	2	3	4	5 หรือมากกว่า
จำนวนชั่วโมง(f)	10	19	31	26	11	3

(X^2 = 8.82 , ยอมรับ Ho)

10.25 ให้ทดสอบว่าข้อมูลในตารางข้างล่างมีการแจกแจงแบบปกติหรือไม่ โดยใช้ α = .05 และกำหนดจำนวนความถี่คาดหมายภายใต้การแจกแจงแบบปกติให้

คะแนน	51 - 60	61 - 70	71 - 80	81 - 90	91 - 100
O_i	3	10	44	50	13
E_i	2	17	50	41	10

(X^2 = 6.98, ยอมรับ Ho)

10.26 บริษัทจัดทำโฆษณาสินค้าได้ประเมินผลการโฆษณาโดยการใช้โทรศัพท์สอบถามว่าได้รับชมโฆษณาชิ้นนั้นหรือไม่ โดยได้ออกอากาศในใช้ช่วงเวลาที่ต่างกัน 3 ครั้ง เมื่อสัปดาห์ก่อน บริษัทเชื่อว่าผู้มีโทรทัศน์จะมีโอกาสได้รับชมโฆษณาแต่ละชิ้นด้วยโอกาส .40 เชื่อว่าการได้รับชมแต่ละครั้งเป็นอิสระกันเพราะใช้เวลาและสถานีต่างกัน ผลการสำรวจผู้ชม 200 คน มีผู้ได้รับชมด้วยจำนวนครั้งต่าง ๆ กัน ดังนี้

จำนวนโฆษณาที่ได้รับชม	0	1	2	3	รวม
ผู้รับชม	46	73	58	23	200

จงใช้ α = .10 ทดสอบว่าจำนวนผู้รับชมมีการแจกแจงแบบทวินามที่มี π = .4
(X^2 = 10.393, ปฏิเสธ Ho)

10.27 จำนวนผู้ประสบอุบัติเหตุกระดูกหักต่อวันของโรงงานพยาบาลหนึ่ง โดยเก็บจากสถิติที่สุ่มมา 300 วัน มีดังนี้

จำนวนคนไข้ต่อวัน	0	1	2	3	4	5	6 ขึ้นไป	รวม
จำนวนวัน	25	45	63	71	48	26	22	300

ถ้าใช้ α = .05 จะสรุปว่าจำนวนคนไข้กระดูกหักของโรงพยาบาลนี้ มีการแจกแจงแบบปัวซองที่มีค่าเฉลี่ย μ = 3 ไหม? (X^2 = 8.35, ยอมรับ Ho)

10.28 ถ้าแบ่งการทำงานของพนักงานดับเพลิงของสถานีดับเพลิงแห่งหนึ่งเป็น 3 ผลัด ๆ ละ 8 ชั่วโมง เชื่อว่ามีโอกาส 30% ที่สัญญาณไฟไหม้จะดัง และมีข้อมูลที่เก็บมาใน 60 วันดังนี้

จำนวนผลัดที่มีสัญญาณเพลิงไหม้	0	1	2	3
จำนวนวัน	16	27	11	6

ให้ใช้ระดับนัยสำคัญ .05 ทดสอบว่า ข้อมูลมีการแจกแจงแบบทวินามที่มี $\pi = .3$

(X^2 = 2.29, ยอมรับ Ho)

10.29 ร้านสรรพสินค้าเก็บสถิติจำนวนลูกค้าที่ต้องการชำระเงินค่าสินค้า ณ เคาน์เตอร์ต่างๆ ในช่วง 5 นาที เขาได้สุ่มเวลา 5 นาที ต่าง ๆ มา 800 ช่วงเวลาพบว่ามีจำนวนลูกค้าชำระเงิน ดังนี้

จำนวนลูกค้าที่เคาน์เตอร์ชำระเงินใน 5 นาที	0	1	2	3	4	5	6ขึ้นไป	รวม
จำนวนครั้ง	36	117	194	167	138	94	54	800

ก. ให้ทดสอบว่าข้อมูลมีการแจกแจงแบบปัวซองที่มี $\mu = 3$ ไหม? $\alpha = .05$

ข. ร้านควรจัดตั้งเคาน์เตอร์ชำระเงินกี่ตัว? (X^2 = 7.369, ยอมรับ Ho, ควรตั้ง 3 จุด)

4. การเปรียบเทียบระหว่าง k สัดส่วน (ของการแจกแจงแบบทวินาม)

ในบทที่ 9 ได้แสดงวิธีทดสอบพารามิเตอร์ π_1 และ π_2 ของการแจกแจงแบบทวินาม โดยใช้การทดสอบแบบ Z แล้ว ถ้าเรามีตัวอย่าง k กลุ่มด้วยขนาด $n_1, n_2, \ldots, n_k$ ตามลำดับ และตัวอย่างเหล่านี้เป็นอิสระกัน และเราต้องการทดสอบความเท่ากันของ k ประชากรที่ตัวอย่างเหล่านี้ถูกสุ่มมา นั่นคือ

Ho : $\pi_1 = \pi_2 = \ldots = \pi_k = \pi$

หรือ Ho : $\pi_j = \pi$ ในเมื่อ $j = 1, 2, \ldots, k$

และ Ha : $\pi_j \neq \pi$

หรือ Ha : π_j ไม่ได้เท่ากันทั้งหมด

ถ้าให้ x_j แทนจำนวน S ในกลุ่มตัวอย่าง n_j ดังนั้น $n_j - x_j$ **คือจำนวน F ในตัวอย่างที่** j และเรามีทั้งหมด k กลุ่มตัวอย่าง ดังนั้น **ข้อมูลที่เก็บมาจะรวบรวมใส่ตารางขนาด 2 × k** ดังนี้

ตัวอย่าง	1	2	...	k	รวม
S	x_1	x_2	...	x_k	x
F	n_1-x_1	n_2-x_2		n_k-x_k	n-x
	n_1	n_2	...	n_k	n

ในเมื่อ

$n_1+n_2+...n_k=n$

$x_1+x_2+...x_k=x$

เมื่อกำหนดระดับนัยสำคัญแล้ว ตัวสถิติที่ใช้ทดสอบคือ

$$X^2 = \sum_{j=1}^{k}\sum_{i=i}^{2}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$ ซึ่งมีการแจกแจงแบบ $\chi^2_{(k-1),\ \alpha}$

ดังนั้น เขตวิกฤตจึงอยู่ด้านขวามือของโค้ง $\chi^2_{(k-1)}$ และเราจะปฏิเสธ Ho เมื่อ $X^2>\chi^2_{(k-1),\ \alpha}$ นั่นคือ เมื่อค่า O_{ij} และ E_{ij} มีความขัดแย้งกันมากจนทำให้ค่า X^2 ที่คำนวณได้เป็นค่าที่โตเกินไป

ตัวอย่าง จำนวนเกษตรกรที่ใช้ปุ๋ย X ใน 3 จังหวัด มีดังนี้

จังหวัด	1	2	3	รวม
ใช้	26	51	33	$110 = R_1$
ไม่ใช้	54	149	127	$330 = R_2$
	$C_1=80$	$C_2=200$	$C_3=160$	$440 = n$

จงทดสอบว่า เกษตรกรใน 3 จังหวัดนิยมปุ๋ย X ไม่แตกต่างกันโดยใช้ $\alpha = .05$

1. $H_o : \pi_1 = \pi_2 = \pi_3 = \pi$ (เกษตรกรใน 3 จังหวัดนิยมใช้ปุ๋ย X ไม่ต่างกัน)
2. H_a : มีอย่างน้อย 1 คู่ที่ต่างกัน (เกษตรกรใน 3 จังหวัดนิยมใช้ปุ๋ย X ต่างกัน)
3. $\alpha = .05$
4. จะปฏิเสธ Ho เมื่อ $X^2 > \chi^2_{(3-1), .05} = 5.99$
5. ในการคำนวณค่าสถิติ $X^2 = \sum_{j-1}^{3} \sum_{i-1}^{2} (O_{ij} - E_{ij})^2 / E_{ij}$

O_{ij} คือข้อมูลที่เก็บมาในตารางขนาด 2 × 3 นั้น ส่วนค่า E_{ij} ต้องคำนวณใหม่โดยสมมุติว่า Ho เป็นจริง $H_o : \pi_1 = \pi_2 = \pi_3 = \pi$ คือสัดส่วนผู้นิยมใช้ปุ๋ย X ในจังหวัดทั้ง 3 เท่ากัน แต่เราไม่ทราบค่า π จึงต้องประมาณจากข้อมูลที่เก็บมาจะได้ $\hat{\pi} = \frac{110}{440} = .25$ นั่นคือ เราคาดว่า จะมี 25% ของเกษตรกรในทุกจังหวัดนิยมใช้ปุ๋ย X ส่วนที่เหลือ $(1 - \hat{\pi}) = .75$ หรือ 75% จะนิยมใช้ปุ๋ยชนิดอื่น ๆ เมื่อเราทราบความน่าจะเป็นของผู้นิยมและไม่นิยมใช้ของทุกจังหวัดแล้ว เราก็หาจำนวนผู้ที่ใช้หรือไม่ใช้ได้ เช่น ของจังหวัดที่ 1 เราสุ่มมาทั้งหมด 80 คน ถ้า Ho เป็นจริงเราคาดว่า

จะมีผู้ใช้ปุ๋ย X = .25(80) = 20 คน ซึ่งมาจาก $\frac{110}{440} \times 80$

ผู้ไม่ใช้ปุ๋ย X = .75(80) = 60 คน ซึ่งมาจาก $\frac{330}{440} \times 80$

ขอให้สังเกตอีกที่ว่า

จำนวนคาดหมายผู้ใช้ปุ๋ย X ในจังหวัดที่ 1 = $E_{11} = \frac{110 \times 80}{440} = \frac{R_1 \times C_1}{n}$

จำนวนคาดหมายผู้ไม่ใช้ปุ๋ย X ในจังหวัดที่ 1 = $E_{21} = \frac{330 \times 80}{440} = \frac{R_2 \times C_1}{n}$

เราจึงสรุปเป็นสูตรทั่วไปเพื่อหาจำนวนคาดหมาย $\boxed{E_{ij} = \frac{R_i \times C_j}{n}}$

ต่อไปเราจะตรวจดูจำนวน df เราจะเห็นว่าในจังหวัดที่ 1 เราสุ่มตัวอย่างมา 80 คน เมื่อเราหาจำนวนคาดหมายของผู้ใช้ปุ๋ย X = 20 คน เราก็จะทราบทันทีว่าจำนวนผู้ไม่ใช้ = 80 - 20 = 60 คน และเป็นจริงเช่นนี้ สำหรับจังหวัดอื่น ๆ ด้วย ดังนั้น ในแต่ละจังหวัดจะมีจำนวนที่เป็นอิสระเพียง 1 จำนวน เพราะเมื่อเราหาค่าคาดหมายของจำนวน S หรือ F เพียง 1 จำนวนแล้ว เหตุการณ์ที่เหลือจะถูกกำหนดทันทีว่าคือส่วนที่เหลือ และเนื่องจากเรามีทั้งหมด k กลุ่ม จำนวน df = k แต่เนื่องจากการหาค่า E_{ij} นั้น เราทราบผลรวม คือค่าที่มุมตารางหมดแล้ว คือเราทราบ n_1, n_2, n_3 เราทราบผลรวม

ของ S คือ x และผลรวมของ F คือ n - x ดังนั้น เมื่อเราหาค่าคาดหมายของ S ได้ 2 จังหวัดแล้ว ของจังหวัดสุดท้ายคือจังหวัดที่ 3 เราก็ไม่ต้องหาเพราะจะคือส่วนที่เหลือจากยอดรวมนั้นเอง ดังนั้น df จึงเหลือ = k - 1 ดังนั้นตาราง E_{ij} จะมีค่าต่าง ๆ ดังนี้

	1	2	3	
ใช้	✓	✓	0	110
ไม่ใช้	0	0	0	330
	80	200	160	440

✓ คือค่าที่หาได้โดยอิสระ

0 คือค่าที่หาได้โดยไม่เป็นอิสระ

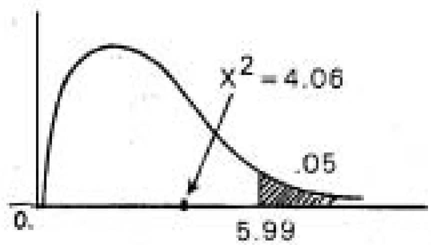
	1	2	3	
ใช้	20	50	40	110
ไม่ใช้	60	150	120	330
	80	200	160	440

ตารางแสดงค่า E_{ij}

$$X^2 = \frac{(26-20)^2}{20} + \frac{(51-50)^2}{50} + \ldots\ldots\ldots\ldots + \frac{(127-120)^2}{120} = 4.06$$

6. ค่าสถิติ $X^2 = 4.06$ ไม่อยู่ในเขตวิกฤต เราจึงยอมรับ-Ho และสรุปว่าสัดส่วนที่แท้จริงของผู้นิยมใช้ปุ๋ย x ใน 3 จังหวัดนั้น ไม่มีความแตกต่างอย่างมีนัยสำคัญ

(p –value > .05 จึงไม่ปฏิเสธ Ho)



ข้อสังเกต มีอยู่จุดหนึ่งที่นักศึกษามักจะเข้าใจสับสนระหว่าง

1. การทดสอบความแตกต่างของ k สัดส่วน (จากประชากรแบบพหุนาม 1 ประชากร)
2. การทดสอบความเป็นอิสระของตารางคอนทินเจนซี
3. การทดสอบความแตกต่างของ k สัดส่วน (จากประชากรแบบทวินาม k ประชากร)

ทั้ง 3 กรณีนี้ จะใช้ตัวสถิติเดียวกันคือ $X^2 = \Sigma\frac{(O-E)^2}{E}$ แต่จุดประสงค์และสมมติฐานต่าง

กัน ในข้อ 1 คือการใช้ X^2 ทดสอบสารรูปสนิทดี คือ ทดสอบว่า ข้อมูลมาจากประชากรแบบพหุนามที่มีพารามิเตอร์ π_{io} ตามที่อ้างไว้ใน Ho หรือไม่ ข้อมูลที่เก็บมามีเพียงตัวอย่างเดียวด้วยขนาดตัวอย่าง n จำนวน และมีลักษณะต่าง ๆ มากกว่า 2 ลักษณะขึ้นไป จึงเรียกว่าพหุนาม เช่น การทดสอบความสมดุลย์ของลูกเต๋า การทดสอบลักษณะต่าง ๆ ของถั่วลูกผสม

สมมติฐานว่างเปล่าและสมมุติฐานรองคือ

Ho : $\pi_1 : \pi_2 : \ldots : \pi_k = \pi_{10} : \pi_{20} : \ldots : \pi_{ko}$

Ha : $\pi_1 : \pi_2 \ldots : \pi_k \neq \pi_{10} : \pi_{20} : \ldots : \pi_{ko}$

ดังนั้น df ที่ใช้เปิดตารางคือ k - 1 ในเมื่อ k = จำนวนลักษณะ ข้อ 2 คือ การแจกแจงแบบพหุนาม 2 ประชากร ประชากรที่ 1 มี r ลักษณะ ประชากรที่ 2 มี c ลักษณะ จึงเก็บข้อมูลจำแนกในตารางคอนทินเจนซีขนาด rxc และเป็นการทดสอบความเป็นอิสระของ 2 ประชากรนี้ ส่วนจำนวน df = (r - 1)(c - 1) จะอธิบายได้โดยใช้ตารางที่ 10.2 ซึ่งสมมุติว่าเรามีตารางขนาด 3 x 4 คอนทินเจนซีดังนี้

คอลัมน์

	1	2	3	4	
แถว 1	✓	✓	✓	0	R_1
แถว 2	✓	✓	✓	0	R_2
แถว 3	0	0	0	*	R_3
	C_1	C_2	C_3	C_4	n

ตารางที่ 10.2 แสดงการหา df ของตารางขนาด 3 x 4

✓ คือค่าที่สามารถคำนวณได้โดยอิสระ

0, * คือค่าที่ไม่สามารถคำนวณได้โดยอิสระ

ส่วนการทดสอบในข้อ 3 มีความคล้ายคลึงกับข้อ 2 มาก ดังจะเห็นได้ว่าข้อมูลที่เก็บมาก็คือ ตารางคอนทินเจนซีขนาด 2 x k การหาค่า E_{ij} ก็ใช้สูตรเดียวกัน คือ $E_{ij} = \frac{R_i X C_j}{n}$ (แต่การหา E_i ในข้อ 1 = $n\pi_{io}$ ไม่ใช่ E_{ij} เพราะ j = 1 คือมีประชากรเดียว) ดังนั้นค่าสถิติ X^2 ที่คำนวณได้จะเท่ากันแม้การหาค่าเปิดตารางก็จะได้ค่าเดียวกัน คือการทดสอบความเป็นอิสระ df = (r − 1)(c − 1) = (2 - 1)(k - 1) = k - 1 = df ของการทดสอบ k สัดส่วนของการแจกแจงแบบทวินามข้อแตกต่างก็คือ **จุดประสงค์**ของการทดสอบและ**การสรุปผล** จากตัวอย่างการทดสอบความแตกต่างของความนิยมปุ๋ย X ใน 3 จังหวัดนั้น ถ้าเราเปลี่ยนคำถามใหม่ว่า "ความนิยมปุ๋ย X กับจังหวัดเป็นอิสระกันไหม" เราจะตั้งสมมุติฐาน Ho ว่า ความนิยมปุ๋ย X เป็นอิสระกับจังหวัดที่ใช้ และ Ha ว่า ความนิยมปุ๋ย

X ไม่เป็นอิสระกับจังหวัดที่ใช้ และค่า X^2 ที่คำนวณได้จะได้เท่าเดิมคือ 4.06 และไม่มีนัยสำคัญ เราจะยอมรับ Ho และสรุปว่า ความนิยมปุ๋ย X กับจังหวัดที่ใช้เป็นอิสระกัน คือไม่ต่างกันนั่นเอง เพราะไม่มีผลกระทบจากปัจจัยอื่นที่จะทำให้ความนิยมต่างกัน สำหรับตัวอย่างข้อนี้อาจไม่ชัดแจ้ง ลองสมมุติตัวอย่างใหม่ แต่จะใช้ข้อมูลเดิม คือ สมมุติข้อมูลในตารางนั้นคือ จำนวนผู้ใช้ยาแก้ปวดหัว 3 ชนิด (กลุ่ม) หลังจากให้กินยา 1 สัปดาห์แล้วถามอาการได้ข้อมูลในตารางข้างล่าง ดังนั้นสมมุติฐานการทดสอบความเป็นอิสระคือ

Ho : ผลการรักษาและชนิดของยาเป็นอิสระกัน (ยาทั้ง 3 ชนิดมีผลการรักษาเท่ากัน)

Ha : ผลการรักษา และชนิดของยาไม่เป็นอิสระกัน (ยาทั้ง 3 ชนิดมีผลการรักษาต่างกัน)

	ยา 1	ยา 2	ยา 3	
หาย	26	51	33	110
ไม่หาย	54	149	127	330
	80	200	160	440

คำนวณ $X^2 = 4.06$, ไม่มีนัยสำคัญ จึงสรุปว่าผลการรักษาไม่ขึ้นกับชนิดของยาที่กิน คือกินยาอะไรก็มีโอกาสหายได้เท่ากัน แต่ในทางตรงข้ามหากข้อมูลในตารางมีความขัดแย้งกับค่าคาดหมายมากทำให้ X^2 คำนวณได้เป็นค่าโตเกินไปจนอยู่ในเขตวิกฤต เราจะปฏิเสธ Ho และสรุปว่า ผลการรักษาขึ้นอยู่กับชนิดของยาที่ใช้รักษา ซึ่งผลสรุปนี้ก็สอดคล้องกับการทดสอบ k สัดส่วนว่าสัดส่วนผู้หายจากโรคของยาแก้ปวด 3 ชนิดแตกต่างกันอย่างมีนัยสำคัญ

5. การวิเคราะห์ความแปรปรวนแบบจำแนก 1 ทาง (One-Way Analysis of Variance)

เรารู้จักใช้การทดสอบแบบไคสแควร์ เพื่อทดสอบความแตกต่างของสัดส่วนจากตัวอย่าง 2 กลุ่มขึ้นไป เพื่อใช้อนุมานว่าตัวอย่างเหล่านั้นมาจากประชากรที่มีสัดส่วนความสำเร็จเหมือนกันหรือไม่ ในหัวข้อนี้ เราจะศึกษาเทคนิคอันหนึ่งซึ่งเรียกว่าการวิเคราะห์ความแปรปรวน หรือ Analysis of Variance ซึ่งมักใช้คำย่อว่า ANOVA ซึ่งจะช่วยให้เราตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยตั้งแต่ 2 กลุ่มตัวอย่างขึ้นไปว่า ถูกสุ่มมาจากประชากรที่มีค่าเฉลี่ยเหมือนกันหรือไม่

เราใช้การวิเคราะห์ความแปรปรวนเมื่อต้องการเปรียบเทียบจำนวนไมล์เฉลี่ยของรถ 5 ชนิด ทดสอบผลการเรียนรู้ของเด็กเมื่อใช้วิธีสอน 4 ชนิด เปรียบเทียบรายได้ปีแรกของผู้จบสาขาธุรกิจจากมหาวิทยาลัยต่าง ๆ เปรียบเทียบผลผลิตข้าวโพดเมื่อใช้ปุ๋ยต่าง ๆ 3 ชนิด ตัวอย่างเหล่านี้ เป็นการเปรียบเทียบค่าเฉลี่ยมากกว่า 2 กลุ่มตัวอย่างขึ้นไปทั้งสิ้น

ตัวอย่าง ผู้จัดการฝ่ายฝึกอบรมต้องการประเมินผลการฝึกงานพนักงานเข้าใหม่ระหว่างวิธีการที่ใช้ 3 วิธี วิธีที่ 1 ให้พนักงานเข้าใหม่ฝึกงานกับพนักงานเก่าที่มีความชำนาญ วิธีที่ 2 ให้พนักงานเข้าใหม่ทั้งหมดเข้าหลักสูตรฝึกงานโดยไม่มีการปฏิบัติงานในโรงงาน วิธีที่ 3 เป็นการฝึกงานโดยมีการฉายภาพยนต์ประกอบการบรรยาย และใช้โปรแกรมวัสดุอื่น ๆ ช่วยด้วย จากการแบ่งพนักงานเข้าใหม่ 16 คน ให้เข้าฝึกงานด้วยวิธีต่าง ๆ เมื่อจบการฝึกและเข้าปฏิบัติงานแล้ว แต่ละคนมีผลผลิตรายวันในตารางที่ 10.3 ผู้จัดการฝ่ายอบรมต้องการทราบว่าวิธีฝึกงาน 3 วิธี ให้มีประสิทธิภาพแตกต่างกันหรือไม่

ตารางที่ 10.3 แสดงผลผลิตของพนักงาน 16 คน

วิธีที่ 1	วิธีที่ 2	วิธีที่ 3	
		18	
15	22	24	
18	27	19	
19	18	16	$N = n_1 + n_2 + n_3 = 16$
22	21	22	$G = 85 + 105 + 114 = 304$
11	17	15	
85	105	114	
÷5	÷5	÷6	
$17 = \bar{X}_1$	$21 = \bar{X}_2$	$19 = \bar{X}_3$	← ค่าเฉลี่ยตัวอย่าง
$n_1 = 5$	$n_2 = 5$	$n_3 = 6$	← ขนาดตัวอย่าง

การหาค่าเฉลี่ยรวมยอด

ค่าเฉลี่ยรวบยอดหรือ grand mean หรือ $\bar{\bar{X}}$ มีวิธีหา 2 วิธี คือ

1. $\bar{\bar{X}} = G/N = 304/16 = 19$

จะเห็นว่าการหา $\bar{\bar{X}}$ ต้องใช้ข้อมูลทั้งหมดจากการทดลอง

2. ใช้วิธีถ่วงน้ำหนักค่าเฉลี่ยตัวอย่างด้วยขนาดตัวอย่าง

$$\bar{\bar{X}} = (\tfrac{5}{16})\,17 + (\tfrac{5}{16})\,21 + (\tfrac{6}{16})\,19$$
$$= \frac{304}{16} = 19$$

สมมติฐานของการทดสอบ

เหตุที่เราต้องใช้การวิเคราะห์ความแปรปรวนก็เพื่อจะได้ตัดสินว่า ตัวอย่าง 3 กลุ่ม (ผลงานของพนักงานภายใต้การฝึกอบรมแต่ละวิธี) มาจากประชากร (ประชากร คือ จำนวนพนักงานทั้งหมดภายใต้การอบรมแต่ละวิธี) ที่มีค่าเฉลี่ยเดียวกัน เรากำลังทดสอบประสิทธิภาพของวิธีฝึกอบรม 3 วิธี นั่นคือ เรากำลังตรวจสอบว่า $\bar{X}_1 = 17, \bar{X}_2 = 21$ และ $\bar{X}_3 = 19$ ซึ่งเป็นตัวแทนของตัวอย่าง 3 กลุ่ม มาจากประชากรที่มีค่าเฉลี่ยเดียวกันคือ μ ดังนั้น เราจึงมีสมมติฐานสำหรับทดสอบดังนี้

$$Ho : \mu_1 = \mu_2 = \mu_3 = \mu$$

$$Ha : \mu_j \text{ ไม่เท่ากันทั้งหมด} \qquad , j = 1, 2, 3$$

ถ้าเราสรุปจากการทดสอบได้ว่า ค่าเฉลี่ยจากตัวอย่างไม่แตกต่างกันอย่างมีนัยสำคัญ เราก็จะสามารถอนุมานได้ว่า วิธีฝึกอบรมไม่มีอิทธิพลต่อผลผลิตของพนักงาน ในทางตรงข้าม หากเราพบว่า ค่าเฉลี่ยจากตัวอย่างมีความแตกต่างกันมากเกินไป หรือมากเกินการยอมรับว่าเกิดจาก sampling error เราจะอนุมานได้ว่าวิธีฝึกอบรม มีอิทธิพลต่อผลผลิตของพนักงาน และเราจะได้ทำการปรับปรุงวิธีการฝึกอบรมให้เหมาะสมต่อไป

หลักเบื้องต้นของการวิเคราะห์ความแปรปรวน

ก่อนที่เราจะใช้เทคนิคการวิเคราะห์ความแปรปรวน เราจะต้องสมมุติได้ว่า ตัวอย่างทั้งหลายนั้นถูกสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ และทุกประชากรมีความแปรปรวนเดียวกัน คือ σ^2 กล่าวโดยสรุป คือ เรามีข้อสมมุติ 2 ข้อ คือ

ถ้า X_{ij} คือข้อมูลตัวที่ i จากประชากรที่ j , $j = 1, 2, \ldots, k$ และ $i = 1, 2, \ldots, n_j$

1. X_{ij} มีการแจกแจงแบบปกติ

2. $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$

หรือเขียนรวมกันได้ว่า $X_{ij} \sim n(\mu_j, \sigma^2)$

สำหรับข้อสมมุติที่ 1 เรื่องการแจกแจงแบบปกตินั้น ถ้าใช้ตัวอย่างขนาดใหญ่ ข้อสมมุตินี้ก็ไม่จำเป็น ซึ่งเราทราบดีจาก central limit theorem

จากตัวอย่าง เรามีสมมติฐานว่างเปล่าว่าทั้ง 3 ประชากรมีค่าเฉลี่ยเดียวกัน ดังนั้น ถ้า- Ho เป็นจริง เราก็ไม่จำเป็นต้องจำแนกข้อมูลเป็นกลุ่มต่าง ๆ 3 กลุ่ม ดังตารางที่ 10.12 เพราะผลผลิตทั้ง 16 จำนวนก็คือตัวอย่าง 1 กลุ่มที่สุ่มมาจาก 1 ประชากร และประชากรรวบยอดนี้มีความแปรปรวน σ^2

การวิเคราะห์ความแปรปรวนมีหลักการเบื้องต้นว่า จะต้องหาค่าประมาณของความแปรปรวนของประชากรรวบยอดคือ σ^2 ซึ่งจะหาค่าประมาณได้ 2 ค่า คือ $\hat{\sigma}_1^2$ และ $\hat{\sigma}_2^2$ เราจะหา $\hat{\sigma}_1^2$ จาก**ความแปรปรวนระหว่างค่าเฉลี่ยตัวอย่าง 3 จำนวนนั้น** ซึ่งคือ 17, 21 และ 19 ส่วน $\hat{\sigma}_2^2$ จะหาจาก**ความผันแปรภายในกลุ่มตัวอย่างทั้ง 3 กลุ่ม** นั่นคือ (15, 18, 19, 22, 11), (22, 27, 18, 21, 17) และ(18, 24, 19, 16, 22, 15) แล้วเราจะนำค่าประมาณของ σ^2 ทั้ง 2 ตัวมาเปรียบเทียบกัน และเพราะว่าทั้งคู่เป็นค่าประมาณของ σ^2 จึงควรจะมีค่าใกล้เคียงกัน ถ้า Ho **เป็นความจริง** แต่ถ้า Ho **เป็นเท็จ** ก็จะให้ค่าประมาณ 2 ตัวนี้ แตกต่างกันมาก ดังนั้น ขั้นตอนของการวิเคราะห์ความแปรปรวนจึงมี 3 ขั้น ดังนี้

1. หาค่าประมาณตัวที่หนึ่งของความแปรปรวนของประชากร จากความแปรปรวนระหว่างค่าเฉลี่ยของตัวอย่าง

2. หาค่าประมาณตัวที่สองของความแปรปรวนของประชากร จากความแปรปรวนภายในกลุ่มตัวอย่าง

3. เปรียบเทียบค่าประมาณทั้ง 2 ตัว ถ้ามีค่าใกล้เคียงกัน จะยอมรับสมมติฐานว่างเปล่า

การหาความแปรปรวนระหว่างค่าเฉลี่ยของตัวอย่าง

เราจะหาค่าประมาณตัวที่หนึ่งของ σ^2 หรือ $\hat{\sigma}_1^2$ โดยหาจากความแปรปรวนระหว่างกลุ่ม

ก่อนอื่นให้เราทบทวนสูตรคำนวณความแปรปรวนจากตัวอย่าง

$$S^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1}$$

แต่เราต้องการหาความแปรปรวนระหว่างค่าเฉลี่ย 3 ตัว ซึ่งแตกต่างจากค่าเฉลี่ยรวบยอด นั่นคือเราให้ $\bar{X}$ แทน X และ $\bar{\bar{X}}$ แทน $\bar{X}$ และ k ถึงจำนวนกลุ่มแทน n เราจะได้สูตรใหม่ ดังนี้

$$S_{\bar{X}}^2 = \frac{\Sigma(\bar{X} - \bar{\bar{X}})^2}{k - 1}$$

และในบทที่ 7 เราทราบว่า

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} \quad \text{หรือ} \quad \sigma^2_{\bar{x}} = \sigma^2/n$$

ดังนั้น $\sigma^2 = \sigma^2_{\bar{x}} \times n$

และ $\hat{\sigma}^2 = \hat{\sigma}^2_{\bar{x}} \times n$

หรือ $\hat{\sigma}^2 = S^2_{\bar{x}} \times n = \dfrac{\Sigma n(\bar{X} - \bar{\bar{X}})^2}{k - 1}$

มีปัญหาว่า n แทนขนาดตัวอย่างของกลุ่มใด ดังนั้น จึงเขียนสูตรให้ชัดเจน ดังนี้

$$\hat{\sigma}^2 = \Sigma_{j=1}^{k} n_j \frac{(\bar{x}_j - \bar{\bar{x}})^2}{k - 1} \qquad \text{.....(10.1)}$$

ในเมื่อ

$\hat{\sigma}^2$ = ค่าประมาณตัวแรกของความแปรปรวนของประชากร (σ^2) ซึ่งหาจากความแปรปรวนระหว่างค่าเฉลี่ย (ความแปรปรวนระหว่างกลุ่ม)

n_j = ขนาดตัวอย่างของกลุ่มที่ j , j = 1, 2, ..., k

$\bar{X}_j$ = ค่าเฉลี่ยของตัวอย่างกลุ่มที่ j

$\bar{\bar{X}}$ = ค่าเฉลี่ยรวมยอด

k = จำนวนกลุ่มตัวอย่าง

เราจะใช้สมการที่ 10.1 หาค่าความแปรปรวนระหว่างกลุ่ม ดังนี้

ตารางที่ 10.4 แสดงการหาค่าความแปรปรวนระหว่างกลุ่ม

n_j	$\bar{X}_j$	$\bar{\bar{X}}$	$\bar{X}_j - \bar{\bar{X}}$	$(\bar{X}_j - \bar{\bar{X}})^2$	$n_j(\bar{X}_j - \bar{\bar{X}})^2$
5	17	19	17 - 19 = -2	$(-2)^2 = 4$	5 × 4 = 20
5	21	19	21 - 19 = 2	$(2)^2 = 4$	5 × 4 = 20
6	19	19	19 - 19 = 0	$(0)^2 = 0$	6 × 0 = 0

$$\hat{\sigma}^2_1 = \frac{\Sigma n_j(\bar{X}_j - \bar{\bar{X}})^2}{k - 1} = \frac{40}{3 - 1} = \frac{40}{2}$$

$\Sigma n_j(\bar{X}_j - \bar{\bar{X}})^2 = 40$

$= 20 \leftarrow$ ความแปรปรวนระหว่างกลุ่ม

การคำนวณความแปรปรวนภายในกลุ่ม

ต่อไปเราจะหาค่าประมาณตัวที่สองของ σ^2 โดยหาจากความแปรปรวนภายในกลุ่มตัวอย่าง โดยใช้สูตร $S^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1}$ เราจะได้ความแปรปรวนภายในกลุ่มตัวอย่าง 3 ตัว คือ S_1^2, S_2^2 และ S_3^2 ตามลำดับ เนื่องจากเราได้มีข้อสมมุติว่าประชากรทั้ง 3 กลุ่มมีความแปรปรวนเดียวกัน ดังนั้น เราจะใช้ค่าใดใน 3 ค่านี้เป็นค่าประมาณของ σ^2 ได้ทั้งสิ้น แต่เรามีวิธีการทางสถิติที่ดีกว่า คือ วิธีการใช้ค่าเฉลี่ยแบบถ่วงน้ำหนักของ 3 ตัวนี้ เป็นค่าประมาณตัวที่ 2 ของ σ^2 เราจึงได้สูตรประมาณค่าอันที่ 2 ดังนี้

$$\hat{\sigma}_2^2 = \frac{\Sigma(n_j - 1)S_j^2}{N - k} \qquad \text{.........10.2}$$

ซึ่งสูตรนี้ มาจาก

$$\hat{\sigma}_2^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + (n_3 - 1)S_3^2}{(n_1 - 1) + (n_2 - 1) + (n_3 - 1)} = S_p^2$$

$$\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + (n_3 - 1)S_3^2}{n_1 + n_2 + n_3 - 3}$$

$$= \frac{\Sigma(n_j - 1)S_j^2}{N - k} \qquad \Sigma n_j = N \quad k = \text{จำนวนกลุ่ม}$$

เราจะหาค่า $\hat{\sigma}_2^2$ ในตารางที่ 10.5 ดังนี้

ตารางที่ 10.5 แสดงการหาความแปรปรวนภายในกลุ่ม

วิธีฝึกอบรมที่ 1		วิธีฝึกอบรมที่ 2		วิธีฝึกอบรมที่ 3	
ค่าเฉลี่ย : $\bar{X}_1 = 17$		ค่าเฉลี่ย : $\bar{X}_2 = 21$		ค่าเฉลี่ย : $\bar{X}_3 = 19$	
$X - \bar{X}_1$	$(X - \bar{X}_1)^2$	$X - \bar{X}_2$	$(X - \bar{X}_2)^2$	$X - \bar{X}_3$	$(X - \bar{X}_3)^2$
$15 - 17 = -2$	$(-2)^2 = 4$	$22 - 21 = 1$	$(1)^2 = 1$	$18 - 19 = -1$	$(-1)^2 = 1$
$18 - 17 = 1$	$(1)^2 = 1$	$27 - 21 = 6$	$(6)^2 = 36$	$24 - 19 = 5$	$(5)^2 = 25$
$19 - 17 = 2$	$(2)^2 = 4$	$18 - 21 = -3$	$(-3)^2 = 9$	$19 - 19 = 0$	$(0)^2 = 0$
$22 - 17 = 5$	$(5)^2 = 25$	$21 - 21 = 0$	$(0)^2 = 0$	$22 - 19 = -3$	$(-3)^2 = 9$
$11 - 17 = -6$	$(-6)^2 = 36$	$17 - 21 = -4$	$(-4)^2 = 16$	$22 - 19 = 3$	$(3)^2 = 9$
$\Sigma(X - \bar{X}_1)^2$	$= 70$	$\Sigma(X - \bar{X}_2)^2$	$= 62$	$15 - 19 = -4$	$(-4)^2 = 16$
$\frac{\Sigma(X - \bar{X}_1)^2}{n_1 - 1}$	$= \frac{70}{5-1}$	$\frac{\Sigma(X - \bar{X}_2)^2}{n_2 - 1}$	$= \frac{62}{5-1}$	$\Sigma(X - \bar{X}_3)^2$	$= 60$
	$= \frac{70}{4}$		$= \frac{62}{4}$	$\frac{\Sigma(X - \bar{X}_3)^2}{n_3 - 1}$	$= \frac{60}{6-1}$
S_1^2	$= 17.5$	S_2^2	$= 15.5$		$= \frac{60}{5}$
				S_3^2	$= 12.0$

$$\hat{\sigma}_2^2 = \frac{\Sigma(n_j - 1)S_j^2}{N - k} = \frac{4(17.5) + 4(15.5) + 5(12.0)}{16 - 3}$$

$$= \frac{192}{13} = 14.769$$

การทดสอบแบบ F

เมื่อได้ค่าประมาณ 2 ตัวของ σ^2 เราจะนำมาเปรียบเทียบกัน และเรียกว่า อัตราส่วน-F หรือ F ratio ดังนี้

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{\text{ความแปรปรวนระหว่างกลุ่ม}}{\text{ความแปรปรวนภายในกลุ่ม}}$$

$$= \frac{20}{14.769}$$

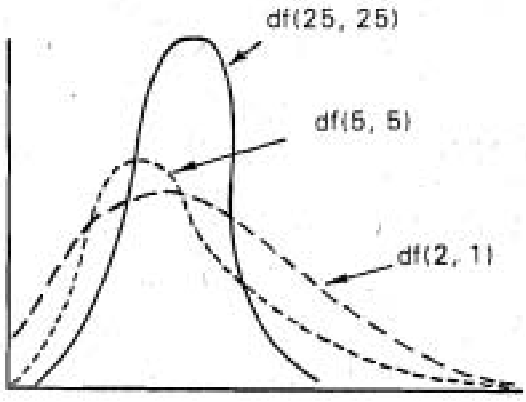
$$= 1.354 \longleftarrow \text{F ratio}$$

เราจะอธิบายค่าสถิติ F อย่างไรดี? ขั้นแรกเราต้องดู df ของตัวตั้งและตัวหารของ F ให้ ν_1 = df ของตัวตั้ง = k - 1 = 3 - 1 = 2 และ ν_2 คือ df ของตัวหาร = N - k = 16 - 3 = 13 ขั้นต่อไปคือ การตรวจสอบตัวหารของ F ว่า หากสมมุติฐานว่างเปล่าที่ตั้งไว้เป็นความจริงหรือไม่ก็ตาม $\hat{\sigma}_2^2$ จะยังคงทำหน้าที่เป็นตัวประมาณค่าที่ดีของ σ^2 ส่วนตัวตั้งของ F คือ $\hat{\sigma}_1^2$ นั้น จะเป็นตัวประมาณค่าที่ดีของ σ^2 ภายใต้ข้อจำกัดว่า Ho ต้องเป็นความจริง นั่นคือวิธีทั้ง 3 มีประสิทธิภาพเท่ากัน ดังนั้น ผลที่ตามมาคืออัตราส่วนของ $\hat{\sigma}_1^2\ \hat{\sigma}_2$ หรือ F ratio จะมีค่าใกล้เคียงหนึ่ง เนื่องจาก $\hat{\sigma}_1^2$ และ $\hat{\sigma}_2^2$ มีค่าใกล้เคียงกัน และเราจะยอมรับว่า Ho เป็นจริงแต่ในทางตรงข้าม หากอัตราส่วน F มีค่าโต ซึ่งจะเป็นผลจากความแปรปรวนระหว่างกลุ่มสูง แสดงว่าตัวอย่างเหล่านั้นไม่ได้มาจากประชากรเดียวกัน หรือ Ho เป็นเท็จ เราจึงมีแนวโน้มที่จะปฏิเสธ Ho สำหรับค่า**โตเกินไป**ของ F ปัญหาต่อไปคือเราจะใช้อะไรเป็นเครื่องวัดว่า F **มีค่าโตเกินไป** หรือไม่ คำตอบก็คือ เราจะต้องนำค่า F ที่คำนวณได้เทียบกับค่าทฤษฎีของมันตามระดับนัยสำคัญที่เรากำหนดไว้ นั่นคือต้องเปิดตารางการแจกแจงแบบ F เทียบดู

ตารางการแจกแจงแบบ F

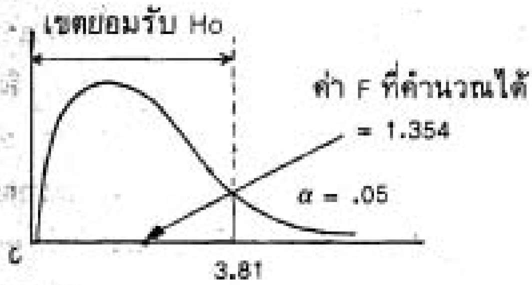
การเปิดตารางการแจกแจงแบบ F ต้องดูเลือกระดับนัยสำคัญ α ก่อนเช่น α = .05, α = .01, α = .025 เป็นต้น การหาค่าในตารางต้องใช้ df 2 อัน คือ ν_1 และ ν_2, ν_1 คือ df ของตัวตั้งของ F จะอยู่ทางแนวตั้ง ส่วน ν_2 คือ df ของตัวหารจะอยู่ทางแนวนอนส่วนรูปร่างการแจกแจงแบบ F จะคล้ายกับการแจกแจงแบบ χ^2 คือ เบ้ขวา, ไม่มีค่าลบ เพราะค่าที่เล็กสุดคือ 0 และรูปร่างจะเปลี่ยนไปตาม df ดังรูปที่ 10.2 จากตัวอย่าง $\nu_1 = 2, \nu_2 = 13$ ถ้าใช้ $\alpha = .05$ เมื่อเปิดตารางที่ $\alpha = .05$ คอลัมน์ที่ 2 แถวที่ 13 จะได้ค่า 3.81 คือค่าที่แบ่งโค้งการแจกแจงเป็น 2 ส่วน โดยให้ส่วนปลายหางมีพื้นที่ 5% ดังรูปที่ 10.3

รูปที่ 10.2
แสดงการแจกแจงของ F ซึ่งมี df ν_1, ν_2 ต่างๆ



การสรุปผล

ค่าสถิติ F = 1.354 ไม่อยู่ในเขตวิกฤต จึงยังปฏิเสธ Ho ไม่ได้ จึงสรุปว่า ไม่มีความแตกต่างอย่างมีนัยสำคัญระหว่างวิธีฝึก 3 วิธี ต่อผลผลิตของพนักงาน



รูปที่ 10.3

(p –value > .05 จึงไม่ปฏิเสธ Ho) แสดงเขตวิกฤตของ $f_{2,\,13}$

สรุปขั้นตอนการทดสอบความแตกต่างของ k ค่าเฉลี่ยดังนี้

1. Ho : $\mu_1 = \mu_2 = ... = \mu_k = \mu$ (ภายใต้ข้อสมมุติว่า $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2$)

2. Ha : μ_j ไม่เท่ากันทั้งหมด หรือ
   Ha : มีค่าเฉลี่ยอย่างน้อย 2 กลุ่ม ที่ต่างกัน

3. กำหนดระดับนัยสำคัญ α

4. เขตวิกฤตอยู่ด้านขวามือของโค้ง f ที่มี $\nu_1 = k - 1$, $\nu_2 = N - k$ นั่นคือจะปฏิเสธ Ho เมื่อ $F > f_{(k-1),\,(N-k),\,\alpha}$

5. คำนวณค่าสถิติ

$$F = \frac{\text{ความแปรปรวนระหว่างกลุ่ม}}{\text{ความแปรปรวนภายในกลุ่ม}}$$

$$= \frac{\Sigma^k n_j(\bar{X}_j - \bar{\bar{X}})^2/(k-1)}{\Sigma^k (n_j - 1)S_j^2/(N-k)}$$

6. ปฏิเสธ Ho ถ้าค่า F อยู่ในเขตวิกฤต

การสร้างตารางวิเคราะห์ความแปรปรวน

การหาค่า F ตามสูตรข้างต้น เป็นการแสดงที่มาของค่าสถิติ F เรายังมีวิธีคำนวณที่ง่ายกว่า โดยใช้ตารางวิเคราะห์ความแปรปรวน

ก่อนอื่นต้องรวบรวมข้อมูลโดยแยกเป็นกลุ่ม หาผลรวม ค่าเฉลี่ยของแต่ละกลุ่ม

ดังแสดงในตารางที่ 10.6 ดังนี้

ตารางที่ 10.6 แสดงข้อมูลจากตัวอย่างสำหรับทำตารางวิเคราะห์ความแปรปรวน

	ตัวอย่าง 1	ตัวอย่าง 2...	ตัวอย่าง k
	X_{11}	X_{12}	X_{1k}
	X_{21}	X_{22}	X_{2k}
	.	.	.
	.	.	.
	.	.	.
	$X_{n_1 1}$	$X_{n_2 1}$	$X_{n_k k}$
ขนาดตัวอย่าง (n_j)	n_1	n_2...	n_k
ผลรวมของตัวอย่าง (T_j)	T_1	T_2...	T_k
จำนวนข้อมูลทั้งหมด :	$n_1 + n_2 + \ldots + n_k = \Sigma n_j = N$		
ผลรวมของข้อมูลทั้งหมด :	$T_1 + T_2 + \ldots + T_k = \Sigma T_j = G$ = Gramd Total		

$$T_1 = \Sigma_{i=1}^{n} X_{i1} = X_{11} + X_{21} + \ldots + X_{n_1 1}$$

$$G = \Sigma_j^k T_j = \Sigma_j^k \Sigma_{i=1}^{n_j} X_{ij} \qquad \text{(แทนค่า } T_j\text{)}$$

จาก

$$F = \hat{\sigma}_1^2 / \hat{\sigma}_2^2$$

ให้ $\hat{\sigma}_1^2$ = SSA/(k - 1), SSA = SUM of squares among groups

$$SSA = \Sigma n_j (\bar{X}_j - \bar{\bar{X}})^2 \quad = \quad \text{(สูตรตามนิยาม)}$$

$$= \Sigma_{j=1}^{k} \frac{T_j^2}{n_j} - \frac{G^2}{N} \qquad (10.3)$$

สมการที่ 10.3 เป็นสมการคำนวณ SSA โดยใช้เครื่องคำนวณ และพึงสังเกตจาก $\hat{\sigma}_1^2$ ว่า SSA มี df = k - 1, ค่า G^2/N เรียกว่า correction factor หรือ CF

ส่วน $\hat{\sigma}_2^2$ = SSE/(N - k) ในเมื่อ SSE = Sum of squares within groups หรือ sum of squares error

มีวิธีคำนวณดังนี้

$$SSE = \Sigma^k (n_j - 1)S_j^2 \quad \text{(สูตรนิยามของ SSE)}$$

$$\text{แต่ } S_j^2 = \Sigma^{n_j}(X_{ij} - \bar{X}_j)^2/(n_j - 1)$$

แทนค่า S_j^2 จะได้

$$SSE = \Sigma^k (n_j - 1)(\Sigma_i^{n_j}(X_{ij} - \bar{X}_j)^2/(n_j - 1))$$

$$= \sum_j^k \sum_i^{n_j} (X_{ij} - \bar{X}_j)^2 \quad \text{(สูตรนิยามของ SSE)}$$

$$= \sum_j^k \sum_i^{n_j} X_{ij}^2 - \Sigma_j^k \frac{T^2}{n_j} \qquad (10.4)$$

สมการที่ 10.4 เป็นวิธีหา SSE โดยใช้เครื่องคำนวณ

และให้ SST = SUM of squares total

$$SST = SSA + SSE$$

$$\text{และ } SST = \sum_j^k \sum_i^{n_j} (X_{ij} - \bar{\bar{X}})^2 \text{ และมี df } = N - 1$$

$$= \sum_j^k \sum_i^{n_j} X_{ij}^2 - G^2/N$$

$$= \sum_j^k \sum_i^{n_j} X_{ij}^2 - CF \qquad (10.5)$$

สมการที่ 10.5 แสดงการหา SST ด้วยเครื่องคำนวณ

ค่า SS ทั้งหลาย และ df เมื่อจัดใส่ตารางวิเคราะห์ความแปรปรวน โดยจำแนกตามที่มา คือ ช่องที่ 1, 2, 3 ได้ดังนี้

ตารางที่ 10.7 ตารางวิเคราะห์ความแปรปรวน หรือ ANOVA

ที่มาของความผันแปร	Sum of Squares (SS)	Degrees of Freedom (df)	ความแปรปรวน Mean Suquare (MS)	F ratio
ระหว่างกลุ่ม (A)	$SSA = \Sigma_j^k \frac{T_j^2}{n_j} - \frac{G^2}{N}$	$\nu_1 = k - 1$	$MSA = S_A^2 = \frac{SSA}{k-1}$	$\frac{S_A^2}{S_E^2} = \frac{MSA}{MSE}$
ภายในกลุ่ม (E)	$SSW = \Sigma_j^k \Sigma_i^{n_j} X_{ij}^2 - \Sigma_j^k \frac{T_j^2}{n_j}$	$\nu_2 = N - k$	$MSE = S_E^2 = \frac{MSE}{N-k}$	
Total (T)	$SST = \Sigma_j^k \Sigma_i^{n_j} x_{ij}^2 - \frac{G^2}{N}$	$N - 1$		

จากตัวอย่างเดิมจะทำตารางวิเคราะห์ความแปรปรวนได้ดังนี้

วิธีที่ 1		วิธีที่ 2		วิธีที่ 3	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
15	225	22	484	18	324
18	324	27	729	24	576
19	361	18	324	19	361
22	484	21	441	16	256
11	121	17	289	22	484
				15	225
85	1515	105	2,267	114	2,226
T_1	$\Sigma_i^5 x_{i1}^2$	T_2	$\Sigma_i^5 x_{i2}^2$	T_3	$\Sigma_i^6 x_{i3}^2$
$n_1 = 5$		$n_2 = 5$		$n_3 = 6$	

$$SSA = \Sigma_j^k \frac{T_j^2}{n_j} - \frac{G^2}{N},$$

$G = \Sigma T_j = 85 + 105 + 114 = 304$

$N = \Sigma n_j = 5 + 5 + 6 = 16$

$$= \frac{85^2}{5} + \frac{105^2}{5} + \frac{114^2}{6} - \frac{304^2}{16}$$

$= 5816 - 5776 = 40$ และมี $\nu_1 = (k - 1) = 2$

และ

$$SSE = \Sigma_j^k \Sigma_i^{nj} X_{ij}^2 - \Sigma_j^k \frac{T_j^2}{n_j}$$

$$= (15^2 + \ldots + 11^2) + (22^2 + \ldots + 17^2) + (18^2 + \ldots + 15^2) - 5{,}816$$

$$= (1515 + 2267 + 2226) - 5816$$

$= 6008 - 5816 = 192$ และมี $\nu_2 = (N - k) = 13$

ดังนั้น $S_A^2 = SSA/(k - 1) = 40/2 = 20$

$S_E^2 = SSE/(N - k) = 192/13 = 14.769$

และ F ratio $= \dfrac{S_A^2}{S_E^2} = \dfrac{20}{14.769} = 1.354$

ตารางวิเคราะห์ความแปรปรวน (ANOVA)

ที่มาของความผันแปรหรือ SOV	ผลบวกกำลังสอง (SS)	df	ความแปรปรวน (MS)	อัตราส่วน F
ระหว่างกลุ่ม (A)	SSA = 40	2	20	$\frac{20}{14.769}$ = 1.354
ภายในกลุ่ม (E)	SSE = 192	13	14.769	
รวม (T)	SST = 232	15		

หมายเหตุ

บางครั้งเราจะไม่หา SSE จากสูตรคำนวณ แต่จะหาจากความสัมพันธ์กับ SST และSSA ดังนี้

$$SST = SSA + SSE$$

$$SSE = SST - SSA$$

$SST = \Sigma\Sigma X_{ij}^2 - G^2/N$ และมี $df = N - 1$

$$SST = 6008 - 5776$$

$$= 232,\ df = 16 - 1 = 15$$

ดังนั้น

$$SSE = 232 - 40$$

$$= 192$$

$$df_E = df_T - df_A$$

$$= 15 - 2 = 13$$

แบบฝึกหัด

10.30 จงแสดงข้อสมมุติของการวิเคราะห์ความแปรปรวน

10.31 การวิเคราะห์ความแปรปรวนเป็นการทดสอบแบบด้านเดียวหรือ 2 ด้าน?

10.32 สมมุติฐานของการวิเคราะห์ความแปรปรวนมีว่าอย่างไร?

10.33 จัดลูกไก่แบบสุ่มให้กินอาหารสูตรต่าง ๆ 3 สูตร แต่ละสูตรมีลูกไก่ 5 ตัว และเมื่อครบกำหนดการทดลอง ได้วัดน้ำหนักที่เพิ่มขึ้น ดังนี้

สูตร 1	สูตร 2	สูตร 3
4	3	6
4	4	7
7	5	7
7	6	7
8	7	8

(F = 2.3, ยอมรับ Ho)

สมมุติให้น้ำหนักที่เพิ่มขึ้นของลูกไก่มีการแจกแจงแบบปกติ จงใช้ α = .05 ทดสอบว่าสูตรทั้ง 3 ให้น้ำหนักเพิ่มต่างกันหรือไม่?

10.34 โรงงานแห่งหนึ่งมีเครื่องจักรชนิดเดียวกัน 4 เครื่อง แต่ละเครื่องจะมีผู้ควบคุม 1 คน เมื่อสุ่มตัวอย่างผลผลิตมา 5 ชั่วโมง ทำงาน พบจำนวนสินค้าชำรุดในแต่ละชั่วโมงทำงานของแต่ละเครื่อง ดังนี้

เครื่องจักร 1	เครื่องจักร 2	เครื่องจักร 3	เครื่องจักร 4
10	7	2	3
9	7	3	3
9	8	3	6
9	8	3	6
8	5	4	7

สมมุติว่าจำนวนชำรุดรายชั่วโมงมีการแจกแจงแบบปกติ จงใช้ α = .01 ทดสอบว่าจำนวนชำรุดจากเครื่องทั้ง 4 ไม่ต่างกัน (F = 1.26, ยอมรับ Ho)

10.35 เมื่อใช้ปุ๋ย 3 ชนิด ในแปลงปลูกสตอเบอรี่ 3 แปลง โดยแปลงที่ 1 ปลูก 5 ต้น แปลงที่ 2 ปลูก 4 ต้น แปลงที่ 3 ปลูก 6 ต้น แต่ละต้นใส่ปุ๋ยจำนวนเท่ากัน ได้ผลผลิตต่อต้น ดังนี้

ปุ๋ย 1	ปุ๋ย 2	ปุ๋ย 3
3	2	4
4	3	5
4	4	6
5	5	6
7		7
		8

สมมุติว่าผลผลิตมีการแจกแจงแบบปกติ จงทดสอบประสิทธิภาพของปุ๋ยทั้ง 3 โดยใช้ α = .01 และ α = .05 (F = 3.84, ยอมรับ Ho)

10.36 สุ่มหลอดไฟชนิดต่าง ๆ 4 ชนิด มาชนิดละ 3 หลอด ได้ผลรวมอายุการใช้งานของแต่ละชนิด (เป็นชั่วโมง) ดังนี้

$T_1 = 50$ $T_2 = 40$ $T_3 = 40$ $T_4 = 50$

และ $\Sigma\Sigma X^2_{ij} = 2778$ สมมุติว่าอายุการใช้งานของหลอดไฟมีการแจกแจงแบบปกติ ให้ทดสอบว่าอายุการใช้งานโดยเฉลี่ยของ 4 ชนิดนั้นแตกต่างกันหรือไม่ด้วย α = .05 (F = 1.99, ยอมรับ Ho)

10.37 ถ้าแบ่งพนักงานของบริษัทหนึ่งตามกลุ่มอายุ 3 กลุ่ม แต่ละกลุ่มสุ่มมา 4 คน พบว่าผลผลิตเฉลี่ยและความแปรปรวนของแต่ละกลุ่ม มีดังนี้

$\bar{X}_1 = 7.5$ $\bar{X}_2 = 10$ $\bar{X}_3 = 8.75$

$S_1^2 = 3.0$ $S_2^2 = 1.0$ $S_3^2 = 1.25$

สมมุติว่า ผลผลิตมีการแจกแจงแบบปกติ จงทดสอบความแตกต่างของผลผลิตของกลุ่มอายุต่าง ๆ โดยใช้ α = .05 (F = 3.57, ยอมรับ Ho)

10.38 ใช้วิธีส่งเสริมจำนวนขาย 4 วิธี ๆ ละ 1 เดือน ได้ข้อมูลคือจำนวนขายจากร้านตัวอย่าง 5 แห่ง ซึ่งใช้วิธีส่งเสริม 4 วิธี ในเดือนต่าง ๆ ดังนี้

แจกตัวอย่างฟรี	77	86	80	88	84
แถมของ 1 ชิ้น	95	92	88	91	89
ลดราคา	72	77	68	82	75
ให้ส่วนลดทางไปรษณีย์	80	84	79	70	82

ก) จงหาจำนวนขายโดยเฉลี่ยของการส่งเสริมแต่ละวิธีและหาค่าเฉลี่ยรวบยอด

ข) จงคำนวณความแปรปรวนระหว่างค่าเฉลี่ยของตัวอย่าง

ค) จงหาค่าประมาณของ σ^2 จากความแปรปรวนระหว่างกลุ่ม

ง) จงหาค่าประมาณของ σ^2 จากความแปรปรวนภายในกลุ่ม

จ) คำนวณค่า F และสรุปด้วยระดับนัยสำคัญ 5% (F = 11.24, ปฏิเสธ Ho)

10.39 จำนวนผู้ยื่นคำร้องขอรับเงินประกันใน 1 วัน ของพนักงานบริษัทประกันภัย 5 คน มีดังนี้

พนักงานคนที่ 1	15	17	14	11		
พนักงานคนที่ 2	12	10	13	17	14	
พนักงานคนที่ 3	10	14	13	15	12	
พนักงานคนที่ 4	14	9	7	10	8	7
พนักงานคนที่ 5	13	12	9	14	10	9

จงใช้ระดับนัยสำคัญ .01 ทดสอบว่าจำนวนผู้ร้องขอรับเงินประกันต่อ 1 วัน ของพนักงานทั้ง 5 คน ไม่ต่างกันอย่างมีนัยสำคัญ (F = 2.62, ยอมรับ Ho)

10.40 จากข้อมูลที่กำหนดให้ข้างล่าง 4 กลุ่มตัวอย่าง เราจะสรุปว่ามาจากประชากรที่มีค่าเฉลี่ยเท่ากัน ด้วยระดับนัยสำคัญ 5% ได้ไหม?

ตัวอย่างที่ 1	17	22	25	29	30	
ตัวอย่างที่ 2	29	18	20	19	30	21
ตัวอย่างที่ 3	13	14	20	18	27	16
ตัวอย่างที่ 4	21	28	20	22	18	

(F = 1.80, ยอมรับ Ho)

10.41 ผู้จัดการโรงงานประดิษฐ์นาฬิกาต้องการทราบว่า เมื่อใช้อัตราความเร็วของสายพานลำเลียงวัสดุต่างกันจะมีผลกระทบอัตราสินค้าชำรุดหรือไม่ เขาได้ทดลองใช้ความเร็ว 4 อัตราในช่วงการทำงาน 5 คาบเวลา ๆ ละ 8 ชั่วโมง และได้บันทึกจำนวนสินค้าชำรุดในแต่ละคาบเวลาไว้ ดังนี้

อัตราความเร็ว

1	2	3	4
36	29	31	36
34	34	35	28
37	34	32	34
35	36	33	32
33	32	39	30

ก) จงหาจำนวนชำรุดโดยเฉลี่ยของแต่ละอัตราความเร็ว และหาค่าเฉลี่ยรวบยอด

ข) จงหาความแปรปรวนระหว่างค่าเฉลี่ยของตัวอย่าง

ค) จงหาความแปรปรวนภายในกลุ่มตัวอย่าง

ง) จงคำนวณอัตราส่วน F และสรุปผลที่ α = .05 (F = 1.13, ยอมรับ Ho)

10.42 ร้านอัดขยายรูปต้องการทราบว่า ระหว่างวิธีส่งเสริมการขาย 3 วิธีคือ

1. แถมฟรีฟิลม์ 1 ม้วน สำหรับทุก ๆ ม้วนที่ล้างและอัด เป็นเวลา 6 สัปดาห์
2. ลดราคา 20 บาท ทุก ๆ ม้วนที่ล้าง-อัด เป็นเวลา 6 สัปดาห์ต่อมา
3. คิดราคาตามปกติ 5 สัปดาห์ก่อนใช้วิธีแจกของหรือลดราคา

จำนวนฟิลม์ที่ลูกค้านำมาให้ล้าง-อัด ต่อสัปดาห์เมื่อใช้วิธีต่าง ๆ มีดังนี้

แจกฟรี 1 ม้วน	65	79	73	55	68	74
ลดราคา 20 บาท	60	64	57	75	62	56
คิดราคาปกติ	61	54	74	59	46	

ก) จงหาจำนวนขายโดยเฉลี่ยต่อสัปดาห์ของแต่ละวิธี และหาค่าเฉลี่ยรวบยอด

ข) จงหาความแปรปรวนระหว่างค่าเฉลี่ยตัวอย่าง

ค) จงหาค่าประมาณของความแปรปรวนของประชากรโดยใช้ความแปรปรวนระหว่างกลุ่ม

ง) จงหาความแปรปรวนภายในกลุ่มตัวอย่าง และหาค่าประมาณของความแปรปรวนของประชากรจากความแปรปรวนภายในกลุ่ม

จ) จงคำนวณค่า F และสรุปผลด้วยระดับนัยสำคัญ α = .05

10.43 ผู้จัดการฝ่ายรักษาความปลอดภัยของร้านสรรพสินค้าขนาดใหญ่ตั้งข้อสงสัยว่าจำนวนผู้ลักขโมยสั่งของจากร้านในช่วงเทศกาลปีใหม่น่าจะสูงกว่าช่วงก่อนหรือหลังปีใหม่ เขาได้เก็บสถิติจำนวนผู้ลักขโมยไว้ 6 ปี มีดังนี้

ก่อนเทศกาลปีใหม่	42	36	58	54	37	47
ระหว่างเทศกาลปีใหม่	51	38	45	32	47	46
หลังเทศกาลปีใหม่	37	29	35	42	31	33

จำนวนผู้ลักขโมยใน 3 ช่วงเวลาแตกต่างกันไหม เมื่อใช้ α = .05
(F = 4.11, ปฏิเสธ Ho)

10.44 บริษัทเคมีภัณฑ์ได้ทดลองวิธีกำจัดคราบน้ำมัน 4 วิธี ข้อมูลข้างล่างคือจำนวนพื้นที่ (ตารางเมตร) ซึ่งเปื้อนคราบน้ำมัน และได้ทำความสะอาดภายใน 1 ชั่วโมง จงทดสอบว่าวิธีทำความสะอาด 4 วิธี มีประสิทธิภาพเหมือนกันที่ระดับนัยสำคัญ 5% ไหม?

วิธี 1 :	55	60	58	61	54	
วิธี 2 :	47	53	54	49	52	(F = 11.44, ปฏิเสธ Ho)
วิธี 3 :	63	59	58	64	63	
วิธี 4 :	51	56	54	59	54	

10.45 โรงงานต้องการทราบ "เวลามาตรฐาน" ในการผลิตสินค้า 1 ชิ้น จึงได้สังเกตการทำงานของพนักงาน 5 คน และบันทึกจำนวนสินค้าที่ผลิตได้ใน 1 ชั่วโมง ของแต่ละคนไว้ เขาเลือกเวลาการผลิตแบบสุ่มของแต่คนมา 9 ชั่วโมงการผลิต ดังตัวเลขข้างล่างคือจำนวนสินค้าที่ผลิตได้ต่อ 1 ชั่วโมง ให้ทดสอบโดยใช้ระดับนัยสำคัญ 1% ว่า ผลผลิตต่อชั่วโมงของพนักงาน 5 คน แตกต่างกันหรือไม่?

คนงาน 1	24	11	19	27	15	16	22	32	17	
คนงาน 2	29	35	37	26	45	26	29	35	38	(F = 10.47,
คนงาน 3	30	28	29	32	22	17	23	29	11	ปฏิเสธ Ho)
คนงาน 4	16	14	5	19	21	17	11	26	9	
คนงาน 5	21	16	19	15	16	28	23	29	17	

การเปรียบเทียบระหว่างค่าเฉลี่ยของ 2 กลุ่ม

เมื่อมีค่าเฉลี่ยมากกว่า 2 กลุ่ม มีบ่อยครั้งที่เราต้องการเปรียบระหว่างค่าเฉลี่ยของ 2 วิธีการ เช่น กลุ่มที่ j กับ j′ นั่นคือเราสนใจ

$$\mu_j - \mu_{j'}$$

ผลต่างระหว่างคู่เฉลี่ยดังกล่าวเรียกว่า **การเปรียบเทียบแบบจับคู่** และตัวประมาณค่าที่ไม่เอียงเฉของ $\mu_j - \mu_{j'}$ คือ $\bar{X}_j - \bar{X}_{j'} = \bar{d}$ และโดยที่ $\bar{X}_j$ และ $\bar{X}_{j'}$ เป็นอิสระกัน ดังนั้นความแปรปรวนของ $\bar{d}$ ก็คือผลบวกของความแปรปรวนของ $\bar{X}_j$ และ $\bar{X}_{j'}$ นั่นเอง นั่นคือ

$$s_{\bar{d}}^2 = \frac{MSE}{n_j} + \frac{MSE}{n_{j'}} \text{ หรือ } MSE\left(\frac{1}{n_j} + \frac{1}{n_{j'}}\right)$$

ดังนั้น เราจึงสร้างช่วงเชื่อมั่นและทดสอบสมมุติฐานของ $\mu_j - \mu_{j'}$ ได้

การสร้างช่วงเชื่อมั่นของ $\mu_j - \mu_{j'}$

$(1-\alpha)$ 100% ช่วงเชื่อมั่นของ $\mu_j - \mu_{j'}$ คือ

$$\bar{d} - t_{\alpha/2,\nu} S_{\bar{d}} < \mu_j - \mu_{j'} < \bar{d} + t_{\alpha/2,\nu} S_{\bar{d}}$$

$$\nu = \text{error df} = N - k, \quad S_{\bar{d}} = \sqrt{MSE\left(\frac{1}{n_j} + \frac{1}{n_{j'}}\right)}$$

การทดสอบการเปรียบเทียบแบบจับคู่

(pairwise coonparison test)

เมื่อเราต้องการทราบว่า $\mu_{j'}$ และ $\mu_{j'}$ แตกต่างกันหรือไม่เราจะมีวิธีการทดสอบดังนี้

1) Ho : $\mu_j - \mu_{j'} = 0$ (หรือ $\mu_j = \mu_j$ นั่นเอง)

2) Ha : $\mu_j - \mu_{j'} \neq 0$ (หรือ $\mu_j \neq \mu_j$ นั่นเอง)

3) กำหนดระดับนัยสำคัญ α

4) จะปฏิเสธ Ho เมื่อ $|\bar{X}_j - \bar{X}_{j'}| > lsd(\alpha)$

โดยที่ $lsd(\alpha) = t_{\alpha/2,\nu} S_{\bar{d}}$

lsd เป็นคำย่อของ least signi ficant difference test

ν คือ df ของ MSE = N - k

$S_{\bar{d}} = \sqrt{MSE\left(\frac{1}{n_j} + \frac{1}{n_{j'}}\right)}$

5) คำนวณค่าสถิติ $|\bar{X}_j - \bar{X}_{j'}|$ และ ค่า $lsd(\alpha)$

โดยที่ $lsd(\alpha) = t_{\alpha/2,\nu}\sqrt{MSE\left(\frac{1}{n_j} + \frac{1}{n_{j'}}\right)}$

6) สรุปผล

ตัวอย่าง จำนวนขายของน้ำอัดลม 4 สี ในตลาดทดลอง มีจำนวนขายโดยเฉลี่ยจากตัวอย่างที่เก็บมา 5 สัปดาห์ และตารางวิเคราะห์ความแปรปรวนดังนี้

	1 ไม่มีสี	2 สีแดง	3 สีส้ม	4 สีเขียว	
จำนวนขายเฉลี่ย	27.32	29.56	26.44	31.46	N = 20
ขนาดตัวอย่าง	$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	$n_4 = 5$	

ตารางวิเคราะห์ความแปรปรวน

ที่มา	SS	df	MS	F-ratio
ระหว่างกลุ่ม (A)	76.85	3	25.62	10.49
Error	39.08	16	2.4425	
รวม	115.93	19		

จากตาราง $f_{.05;\,3,\,16} = 3.24$

สรุปผลได้ว่า จำนวนขายโดยเฉลี่ยของน้ำอัดลมสีต่าง ๆ ไม่เท่ากันทั้งหมด จะเห็นว่าสีแดงและสีเขียวมีจำนวนขายสูงกว่าสีอื่น ถ้าบริษัทต้องการผลิตสีเดียวเพื่อจำหน่ายทั่วประเทศ และบริษัทต้องการทราบว่าจำนวนขายโดยเฉลี่ยที่แท้จริงของสีแดง และเขียว มีความแตกต่างอย่างมีนัยสำคัญหรือไม่ เขาจะมีวิธีการทางสถิติ 2 วิธีที่จะตอบปัญหาได้ คือการสร้างช่วงเชื่อมั่นของ $\mu_4 - \mu_2$ และการทดสอบความแตกต่างของคู่เฉลี่ย μ_4 และ μ_2 ถ้ากำหนดระดับนัยสำคัญ 5% จะได้ 95 % ช่วงเชื่อมั่นของ $\mu_4 - \mu_2$

$$= (\bar{X}_4 - \bar{X}_2) \pm t_{.025,\,16} \sqrt{MSE\left(\frac{1}{5} + \frac{1}{5}\right)}$$

$$= (31.46 - 29.56) \pm 2.120 \sqrt{2.4425\left(\frac{1}{5} + \frac{1}{5}\right)}$$

$$= 1.90 \pm 2.120\,(.988)$$

$$= 1.90 \pm 2.095$$

$$= -0.195, 3.995$$

$$= -0.195 < \mu_4 - \mu_2 < 3.995$$

จะเห็นว่าช่วงเชื่อมั่นรวมค่า 0 ($\mu_4 - \mu_2 = 0$) ด้วย นั่นคือความแตกต่างของ μ_4 และ μ_2 ไม่มีนัยสำคัญ

	$\bar{X}_2$ 29.56	$\bar{X}_3$ 26.44	$\bar{X}_4$ 31.46
$\bar{X}_1$=27.32	2.24*	0.88	4.14*
$\bar{X}_2$=29.56	–	3.12*	1.90*
$\bar{X}_3$=26.44	–	–	5.02*

การทดสอบสมมติฐาน

1. Ho : $\mu_4 - \mu_2 = 0$
2. Ha : $\mu_4 - \mu_2 \neq 0$
3. $\alpha = .05$
4. $lsd(.05) = t_{.025,\,16}\sqrt{MSE\left(\frac{1}{5}+\frac{1}{5}\right)}$

$= 2.120 \sqrt{2.4425\left(\frac{2}{5}\right)}$

$= 2.095$

เราจะปฏิเสธ Ho ถ้า $\bar{X}_4 - \bar{X}_2 > lsd(.05)$

นั่นคือเมื่อ $\bar{X}_4 - \bar{X}_2 > 2.095$

5. $\bar{X}_4 - \bar{X}_2| = 31.46 - 29.56$

$= 1.90$

6. เพราะว่า 1.90 < 2.095 ยังปฏิเสธ Ho ไม่ได้ จึงสรุปว่าจำนวนขายโดยเฉลี่ยของน้ำอัดลมทั้ง 2 สีไม่ต่างกันอย่างมีนัยสำคัญ

> **กรณีต้องการทดสอบนัยสำคัญผลต่างทุกคู่ ควรสร้างตารางแสดงผลต่างค่าเฉลี่ยทุกคู่ เช่นตารางข้างบน จะพบว่ามีผลต่าง 4 คู่ ที่มากกว่า lsd (.05) = 2.095 จึงสรุปว่า $\mu_1 \neq \mu_2$, $\mu_1 \neq \mu_4$, $\mu_2 \neq \mu_3$ และ $\mu_3 \neq \mu_4$**

แบบฝึกหัด

10.46 สำนักงานวิจัยตลาดต้องการทราบว่า เมื่อเขียนจดหมายเชิญชวนให้เป็นสมาชิกวารสารรายสัปดาห์ฉบับหนึ่งด้วยวิธีต่าง ๆ 3 วิธี คือ

1. ใช้ตัวพิมพ์ทั้งฉบับ
2. ใช้ตัวพิมพ์แต่มีลายเซ็นต์ผู้ส่งจดหมาย
3. ใช้ตัวเขียนทั้งฉบับ

เขาเลือกท้องที่ที่มีประชากรขนาดใกล้เคียงกันได้ 12 ท้องที่ จึงทำการส่งจดหมายเชิญชวนไปท้องที่ละ 10,000 ฉบับ โดยใช้วิธีต่าง ๆ 3 วิธีดังกล่าววิธีละ 4 ท้องที่ มีจดหมายตอบรับเป็นสมาชิก ดังนี้

วิธีการ (j)

1	2	3
606	660	671
655	643	724
550	595	639
613	670	762

ก. ความหมายของ Ho : $\mu_1 = \mu_2 = \mu_3$ คือ ไม่มีวิธีใดจาก 3 วิธีที่เหมาะสมใช่หรือไม่? จงอธิบาย

ข. จงทดสอบเปรียบเทียบประสิทธิภาพของวิธีทั้ง 3 โดยใช้ $\alpha = .05$

ค. จากการทดสอบแบบ F จะสรุปว่าจดหมายแบบ 1 และ 2 มีผลต่างกันได้ไหม? จงอธิบาย

(F = 4.43, ปฏิเสธ Ho, ยอมรับ Ha : $\mu_i \neq \mu_j$

10.47 จากข้อ 10.46 ถ้าข้อมูลมีดังนี้

วิธีการ (j)

1	2	3
603	666	729
634	625	687
594	622	711
639	661	693

(F = 17.33, ปฏิเสธ Ho ; ยอมรับ Ha : $\mu_i \neq \mu_j$

ให้ใช้คำถามเหมือนข้อ 10.46

10.48 จากข้อ 10.46

ก) จงสร้างช่วงเชื่อมั่น 95% ของ μ_3 และอธิบายช่วงเชื่อมั่นที่หาได้ (648.58 ; 749.42)

ข) จงสร้างช่วงเชื่อมั่น 95% ของ $(\mu_3-\mu_1)$ และอธิบายผลที่ได้ (21.7 ; 164.3)

ค) เราจะมีความมั่นใจว่าจดหมายแบบที่ 3 ให้ผลดีกว่าแบบที่ 1 ไหม?

10.49 จงใช้ข้อมูลจากข้อ 10.47 และใช้คำถามข้อ 10.48

6. การวิเคราะห์ความแปรปรวนแบบจำแนก 2 ทาง

(Two-Way Analysis of Variance)

การทดสอบแบบ F ในหัวข้อที่ 5 เป็นวิธีการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว คือได้แบ่งความผันแปรทั้งหมดออกเป็น 2 ส่วน คือ ระหว่างกลุ่ม และภายในกลุ่ม ความผันแปรระหว่างกลุ่มเป็นความผันแปรที่เราเห็นชัดเจน บางครั้งเราเรียกว่า factor หรือ treatment ส่วนความผันแปรภายในกลุ่มจะรวมอิทธิพลอื่น ๆ นอกจาก factor ไว้ และมักเรียกว่า ความคลาดเคลื่อนจากการทดลองหรือ experimental error เรียกสั้น ๆ ว่า error ยังมีวิธีวิเคราะห์ความแปรปรวนอีกหลายวิธี แต่จะขอกล่าววิธีหนึ่งซึ่งเรียกว่า การวิเคราะห์ความแปรปรวนแบบจำแนก 2 ทาง โดยวิธีจะได้ข้อมูลจำแนกตามแฟคเตอร์ 2 แฟคเตอร์ เช่น ในระบบการผลิต เราต้องใช้คนงานหลายคน และใช้เครื่องจักรหลาย ๆ เครื่อง เราสามารถทดสอบความแตกต่างของผลผลิตจากคนงานหลาย ๆ คน และสามารถทดสอบความแตกต่างในการผลิตของเครื่องจักรหลาย ๆ เครื่อง ตามวิธีการในหัวข้อที่ 5 แต่ถ้าเราต้องการตรวจสอบพร้อม ๆ กัน ทั้ง 2 อย่าง และในขณะเดียวกัน เราต้องการทราบอิทธิพลร่วมกันของ 2 แฟคเตอร์ ซึ่งเรียกว่า joint effect หรือ interaction effectเพราะคนงานบางคนอาจคุ้นเคยกับเครื่องจักรบางเครื่องเป็นพิเศษ จึงให้ผลงานดีกว่าเมื่อใช้เครื่องอื่น ๆ ตัวอย่างอีกอันหนึ่งคือ การวิจัยตลาด ซึ่งเราจะจำแนกพนักงานขาย 2 ลักษณะ (แฟคเตอร์) คือ จำแนกตามอายุและประสบการณ์ และถ้าเราต้องการตรวจสอบอิทธิพลของอายุ และอิทธิพลของประสบการณ์ต่อจำนวนขายของพนักงาน ในกรณีอาจตรวจพบว่าอายุและประสบการณ์มีอิทธิพลร่วมกัน เพราะในกลุ่มอายุบางช่วงผสมกับประสบการณ์ระดับหนึ่ง อาจทำให้จำนวนขายมากเป็นพิเศษ

เมื่อเราพิจารณา 2 แฟคเตอร์พร้อมกัน เราจะต้องเก็บรวบรวมข้อมูลแสดงในตาราง 2 ทาง ขนาด $r \times k$ ในเมื่อ r คือ จำนวนแถว และ k คือจำนวนคอลัมน์ และเพื่อจะแยกอิทธิพลของ-interaction ออกจาก random error เราจะต้องมีข้อมูลในแต่ละเซลล์ไม่น้อยกว่า 1 ตัว คือต้องมี **"การซ้ำ"** (เซลล์คือ ส่วนตัดของแถวกับคอลัมน์) เพราะถ้าเรามีข้อมูลเพียง 1 ตัว ในแต่ละเซลล์ เราจะไม่สามารถบอกได้ว่า ผลต่างระหว่างเซลล์เป็นสาเหตุจาก random error แต่เพียงอย่างเดียว หรือเป็นผลรวมของ random error และ interaction ข้อมูลของตารางวิเคราะห์ความแปรปรวนแบบ 2 ทาง มีดังนี้

ตารางที่ 10.8 แสดงข้อมูลการวิเคราะห์ความแปรปรวนแบบ 2 ทาง

		คอลัมน์ (k)				รวม
		1	2	...	k	
	1	X_{111} X_{112} ... X_{11n}	X_{121} X_{122} ... X_{12n}	... ...	X_{1k1} X_{1k2} ... X_{1kn}	R_1
แถว (j)	2	X_{211} X_{212} ... X_{21n}	X_{221} X_{222} ... X_{22n}	... ...	X_{2k1} X_{2k2} ... X_{2kn}	R_2
	...	...	...	...	...	
	r	X_{r11} X_{r12} ... X_{r1n}	X_{r21} X_{r22} ... X_{r2n}	... ...	X_{rk1} X_{rk2} ... X_{rkn}	R_r
รวม		C_1	C_2	...	C_k	G

ขอให้สังเกตว่าทุก ๆ เซลล์มีจำนวนซ้ำเท่ากัน คือ n จำนวน
เรามี r แถว และ k คอลัมน์ ดังนั้นข้อมูลที่เก็บมาทั้งหมด
จะมี $N = r \times k \times n$ จำนวน
ให้ X_{jki} แทนข้อมูลในแถว j คอลัมน์ k และตัวที่ i
โดยที่

$j = 1, 2, \ldots, r$

$k = 1, 2, \ldots, K$

$i = 1, 2, \ldots, n$

ให้ SSR แทน SS ของ อิทธิพลด้านแถว

SSC แทน SS ของ อิทธิพลด้านคอลัมน์

SSI แทน SS ของ อิทธิพลร่วมกัน (interaction)

SSE แทน SS ของ error

SST แทน SS ของ Total

เราจะหาค่า SS ของอิทธิพลต่าง ๆ ดังนี้

1 $$SSR = \Sigma_j^r \Sigma_k^K \Sigma_i^n (\bar{X}_j - \bar{\bar{X}})^2$$
$$= \Sigma_j^r nK (\bar{X}_j - \bar{\bar{X}})^2$$
$$= \Sigma_{j=1}^r \frac{R_j^2}{nK} - \frac{G^2}{N}$$

2 $$SSC = \Sigma_j^r \Sigma_k^K \Sigma_i^n (\bar{X}_j - \bar{\bar{X}})^2$$
$$= \Sigma_k^K rn (\bar{X}_k - \bar{\bar{X}})^2$$
$$= \Sigma_{k=1}^K \frac{C_k^2}{rn} - \frac{G^2}{N}$$

3 $$SSI = SST - SSR - SSC - SSE$$

4 $$SSE = \Sigma \Sigma \Sigma (X_{jki} - \bar{X}_{jk})^2$$
$$= \Sigma_j^r \Sigma_k^K \Sigma_i^n X_{jki}^2 - \Sigma_j^r \Sigma_k^K \frac{(\Sigma_i^n X_{jki})^2}{n}$$

5 $$SST = \Sigma_j^r \Sigma_k^K \Sigma_i^n (X_{ijk} - \bar{\bar{X}})^2$$
$$= \Sigma_j^r \Sigma_k^K \Sigma_i^n X_{jki}^2 - \frac{G^2}{N}$$

ในปัจจุบันนี้ ธุรกิจทั่วไปนิยมใช้เครื่องคอมพิวเตอร์ขนาดเล็ก ซึ่งมีโปรแกรมของการวิเคราะห์ความแปรปรวน จึงหมดปัญหาด้านการคำนวณ ในกรณีที่ไม่มีเครื่องคอมพิวเตอร์ เราก็สามารถใช้เครื่องคำนวณแบบตั้งโต๊ะ หรือแม้แต่เครื่องคำนวณขนาดเล็กก็พอ

ตารางที่ 10.9 แสดงการวิเคราะห์ความแปรปรวนแบบจำแนก 2 ทาง

ที่มา	SS	df	MS	F ratio
ระหว่างแถว = R	SSR	r - 1	$\frac{SSR}{r-1}$	$\frac{MSR}{MSE}$
ระหว่างคอลัมน์ = C	SSC	K - 1	$\frac{SSC}{K-1}$	$\frac{MSC}{MSE}$
อิทธิพลร่วมของแถวและคอลัมน์ (Interaction = I)	SSI	(r - 1)(K - 1)	$\frac{SSI}{(r-1)(K-1)}$	$\frac{MSI}{MSE}$
ERRor = E	SSE	rK(n - 1)	$\frac{SSE}{rK(n-1)}$	
ผลรวม = T	SST	rKn - 1		

เมื่อจะทดสอบ Ho : ไม่มีอิทธิพลร่วมกันระหว่าง 2 แฟคเตอร์

ตัวทดสอบคือ $F = \frac{MSI}{MSE}$ และเปรียบเทียบกับค่าในตาราง f ที่

$\nu_1 = (r-1)(k-1)$ และ $\nu_2 = rK(n-1)$

เมื่อจะทดสอบ Ho : ไม่มีอิทธิพลทางด้านแถว

ตัวสถิติทดสอบ คือ $F = \frac{MSR}{MSE}$ และเปรียบเทียบกับค่า f ในตารางที่ $\nu_1 = r-1$ และ $\nu_2 = rK(n-1)$

เมื่อจะทดสอบ Ho : ไม่มีอิทธิพลทางด้านคอลัมน์

ตัวสถิติทดสอบคือ $F = \frac{MSC}{MSE}$ และเปรียบเทียบกับค่า ในตารางที่ $\nu_1 = K-1$ และ $\nu_2 = rk(n-1)$

โปรดสังเกตว่า ตัวสถิติที่ใช้ทดสอบอิทธิพลต่างมี MSE เป็นตัวหารทั้งสิ้น ในกรณีที่ทดสอบแล้ว ไม่มีอิทธิพลร่วมกันระหว่างแถวและคอลัมน์ เราอาจหาตัวหารใหม่ เรียกว่า pooled eror โดยนำ SSI รวมกับ SSE แล้วหารด้วย ผลรวม df ของ SSI และ SSE

ตัวอย่าง โรงงานแห่งหนึ่งใช้เครื่องจักรผลิตสินค้าชนิดต่าง ๆ กัน 5 เครื่อง และมีพนักงานควบคุมเครื่องจักร 4 คน โรงงานต้องการทราบว่า มีความแตกต่างในจำนวนสินค้าชำรุดซึ่งผลิตโดยคนงาน

แต่ละคนหรือไม่ และต้องการทราบจำนวนสินค้าชำรุดจากเครื่องจักรต่าง ๆ มีความแตกต่างหรือไม่ และต้องการทราบว่า มีอิทธิพลร่วมกันระหว่างพนักงานและเครื่องจักรหรือไม่ จึงให้พนักงานทุกคนได้มีโอกาสใช้เครื่องจักรทุกเครื่อง ๆ ละ 2 วัน และบันทึกจำนวนสินค้าชำรุดในแต่ละวันไว้ในตารางที่ 10.10 ดังนี้

ตารางที่ 10.10 แสดงจำนวนสินค้าชำรุดของพนักงาน 4 คน และเครื่องจักร 5 เครื่อง

เครื่องจักร	พนักงาน รวม					
		1	2	3	4	
1		34	36	37	38	
		28	26	35	24	$R_1 = 258$
	รวม	62	62	72	62	
2		34	34	25	28	
		31	22	40	29	$R_2 = 243$
	รวม	65	56	65	57	
3		20	33	30	22	
		19	20	22	12	$R_3 = 178$
	รวม	39	53	52	34	
4		23	36	36	30	
		30	18	14	21	$R_4 = 208$
	รวม	53	54	50	51	
5		29	42	45	32	
		34	34	33	25	$R_5 = 274$
	รวม	63	76	78	57	
รวม		$C_1 = 282$	$C_2 = 301$	$C_3 = 307$	$C_4 = 261$	$G = 1161$

$$r = 5,\ K = 4,\ n = 2,\ N = 5(4)(2) = 40$$

(1) G^2/N $= \quad CF = (1161)^2/40 = 33{,}698.025$ และมี df = 1

(2) SSR $= \Sigma_i^r \frac{R_i^2}{nk} - CF$

$= \frac{258^2}{8} + \frac{243^2}{8} + \frac{178^2}{8} + \frac{208^2}{8} + \frac{274^2}{8} - 33{,}698.025$

$= \frac{275{,}637}{8} - 33{,}698.025 = 756.6$ และมี df = 4

(3) SSC $\quad \Sigma_k^K \frac{C_k^2}{nr} - CF$

$= \frac{282^2 + 301^2 + 307^2 + 261^2}{2(5)} - 33{,}698.025$

$= 33{,}873.5 - 33{,}698.025 = 175.475$ และมี df = 3

(4) SSE $= \Sigma_j^r \Sigma_k^K \Sigma_i^n X_{ijk}^2 - \Sigma_j^r \Sigma_k^K \frac{(\Sigma_i^n X_{jki})^2}{n}$

(4.1) $\Sigma_j^r \Sigma_k^K \Sigma_i^n X_{ijk}^2$ $= (34^2 + 28^2 + 36^2 + \ldots + 32^2 + 25^2)$ รวม 40 จำนวน

$= 35{,}957$ และมี df = 40

(4.2) $\Sigma_j^r \Sigma_k^K \frac{(\Sigma_i^n X_{jki})^2}{n}$ $= \frac{(62^2 + 62^2 + 72^2 + \ldots + 78^2 + 57^2)}{2}$ รวม 20 จำนวน

$= 34{,}822.5$ และมี df = 20

ดังนั้น SSE $= (4.1) - (4.2)$

$= 35{,}957 - 34{,}822.5$

$= 1{,}134.5$ และมี df $= (40 - 20) = 20$

(4) SST $= \Sigma_j^r \Sigma_k^K \Sigma_i^n X_{jki}^2 - G^2/N$

$= (4.1) - (1)$

$= 35{,}957 - 33{,}698.025$

$= 2{,}258.975$ และมี df $= 40 - 1 = 39$

(5) SSI $=$ SST - SSR - SSC - SSE

$= 2{,}258.975 - 756.6 - 175.475 - 1{,}134.5$

$= 192.4$ และมี df $= 39 - 4 - 3 - 20 = 12$

ตารางวิเคราะห์ความแปรปรวน

ที่มา	SS	df	MS	F ratio
R = ระหว่างเครื่องจักร	SSR = 756.6	5 - 1 = 4	$\frac{756.6}{4} = 189.15$	$\frac{189.15}{56.725} = 3.33$
C = ระหว่างพนักงาน	SSC = 175.475	4 - 1 = 3	$\frac{175.475}{3} = 58.49$	$\frac{58.49}{56.725} = 1.03$
I = อิทธิพลร่วม	SSI = 192.4	(5-1)(4-1) = 12	$\frac{192}{12} = 16.03$	$\frac{16.03}{56.725} = 0.28$
E = Error	SSE = 1134.5	5(4)(1) = 20	$\frac{1134.5}{20} = 56.725$	
T = ผลรวม	2258.975	5(4)(2) 1 = 39		

1. Ho : ไม่มีอิทธิพลร่วมกันระหว่างเครื่องจักร และพนักงาน

   F = 0.28, จากตาราง $f^{.05}_{12, 20} = 2.28$

   จึงยอมรับ Ho สรุปว่า ไม่มีอิทธิพลร่วมกัน

2. Ho : ไม่มีอิทธิพลเนื่องจากเครื่องจักร

   F = 3.33, จากตาราง $f^{.05}_{4, 20} = 2.87$

   ค่าคำนวณใหญ่กว่าค่าตารางจึงปฏิเสธ Ho และสรุปว่า มีความแตกต่างอย่างมีนัยสำคัญระหว่างจำนวนเฉลี่ยสินค้าชำรุดซึ่งผลิตโดยเครื่องจักรทั้ง 5 นั้น

3. Ho : ไม่มีอิทธิพลเนื่องจากพนักงาน

   F = 1.03, จากตาราง $f^{.05}_{3, 20} = 3.10$

   ค่าสถิติไม่มีนัยสำคัญเพราะเล็กกว่าค่าตาราง จึงสรุปว่าไม่มีความแตกต่างอย่างมีนัยสำคัญระหว่างสินค้าชำรุดโดยเฉลี่ย ซึ่งผลิตโดยพนักงาน 4 คนนั้น

แบบฝึกหัด

10.50 โรงงานแห่งหนึ่งใช้เครื่องจักรผลิตสินค้า 3 เครื่อง และพนักงานควบคุมเครื่องจักร 3 คน เพื่อจะทดสอบอิทธิพลของเครื่องจักร, อิทธิพลของคนงานและอิทธิพลร่วมกันของเครื่องจักรและคนงาน จึงให้พนักงานทุกคนได้มีโอกาสควบคุมทุกเครื่อง ๆ ละ 2 ชั่วโมง และเก็บข้อมูลคือ จำนวนผลผลิตต่อ 1 ชั่วโมง ดังนี้

พนักงาน	เครื่องจักร		
	1	2	3
1	22	16	13
	18	14	10
2	18	27	16
	14	19	13
3	17	17	14
	14	11	9

F (พนักงาน) = 2.5

F (เครื่องจักร) = 4.32*

F (อิทธิพลร่วม) = 1.95

จงทดสอบอิทธิพลต่าง ๆ ด้วยระดับนัยสำคัญ 5% (สมมุติว่าผลผลิตมีการแจกแจงแบบปกติ)

10.51 งานทดลองอีกอันหนึ่งมี 2 แฟคเตอร์ แฟคเตอร์ด้านแถวมี 4 ระดับ และแฟคเตอร์ด้านคอลัมน์มี 3 ระดับ และมีข้อมูล 5 ตัว ในทุก ๆ เซลล์ซึ่งคือส่วนผสมของแถว j คอลัมน์ k และคำนวณค่า SS ได้ดังนี้

SSR = 315

SSC = 405

SSI = 900

SSE = 960

F (แถว) = 5.25*

F (คอลัมน์) = 10.125*

F (อิทธิพลร่วม) = 75**

จงใช้ระดับนัยสำคัญ 5% ทดสอบอิทธิพลต่าง ๆ

10.52 บริษัทประกันต้องการวิเคราะห์ความแปรปรวนแบบ 2 ทาง เพื่อศึกษาอิทธิพลของ 3 กลุ่มอายุ (คอลัมน์) และอิทธิพลของระดับการศึกษา 4 ระดับ (แถว) และมีข้อมูล 5 ตัว ในทุกเซลล์ jK ได้ข้อมูลเบื้องต้นดังนี้

$$\Sigma_j \Sigma_k \Sigma_i X_{jki}^2 = 4010,\ G^2 / N = 620$$

$$\Sigma_j \Sigma_k (\Sigma_i^n X_{jki})^2 / n = 3050,\ \Sigma_j^r \frac{R_j^2}{nK} = 890,\ \Sigma_{k=1}^K \frac{C_k^2}{nr} = 980$$

จงทดสอบอิทธิพลของแถว, คอลัมน์ และผลร่วมโดยใช้ $\alpha = .01$

F(การศึกษา) = 4.5**, F(อายุ) = 9**, F(อิทธิพลร่วม) = 15.0**

10.53 ปลูกข้าว 3 พันธุ์ โดยใช้ปุ๋ย 3 ชนิด ในแปลงปลูกที่มีสภาพไม่ต่างกัน โดยปลูกส่วนผสมของพันธุ์ข้าว และปุ๋ยชุดละ 2 แปลง ได้ผลผลิต ดังนี้

ปุ๋ย	พันธุ์ข้าว		
	1	2	3
1	10	7	5
	12	8	3
2	13	9	4
	13	10	5
3	12	10	6
	10	8	5

จงทดสอบอิทธิพลของปุ๋ย, พันธุ์ข้าว และอิทธิพลร่วมกัน, $\alpha = .05$

(F(ปุ๋ย) = 3.15, F(พันธุ์ข้าว) = 66.67, F(อิทธิพลร่วม) = 1.13)

7. การอนุมานความแปรปรวนของ 1 ประชากร

(Inference about a population variance)

เราทราบวิธีการสร้างช่วงเชื่อมั่นและทดสอบสมมติฐานสำหรับค่าเฉลี่ย และสัดส่วน จาก 1 ประชากร และ 2 ประชากร ในบทที่ 8 และ 9 แล้ว นอกจากนั้น เรายังรู้จักใช้การทดสอบแบบไคสแควร์ และแบบ F สำหรับอนุมานเมื่อมีค่าเฉลี่ยและสัดส่วนมากกว่า 2 กลุ่มขึ้นไป ยังมีพารามิเตอร์ที่สำคัญอีกตัวหนึ่งคือ ความแปรปรวนของประชากรที่ยังมิได้กล่าวเลย

การประมาณค่าความแปรปรวนของ 1 ประชากร

เราทราบว่าค่าประมาณแบบจุดของ σ^2 คือ s^2 แต่เรายังไม่ทราบการแจกแจงของ s^2 จึงยังสร้างช่วงเชื่อมั่นของ σ^2 ไม่ได้ มีตัวสถิติที่สำคัญคือ

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

จะมีการแจกแจงแบบ χ^2 ที่มี $\nu = (n-1)$

ดังนั้น $$\sigma^2 = \frac{(n-1)s^2}{\chi^2}$$

เราจะได้ช่วงเชื่อมั่นของ σ^2 ดังนี้

$$\sigma_L^2 = \frac{(n-1)s^2}{\chi_U^2} \qquad \text{ขีดจำกัดล่าง}$$

$$\sigma_U^2 = \frac{(n-1)s^2}{\chi_L^2} \qquad \text{ขีดจำกัดบน}$$

ในเมื่อ χ_U^2 และ χ_L^2 คือค่าที่เปิดจากตาราง χ^2 ที่มี df = (n - 1) และทำให้เหลือพื้นที่ด้านขวา และด้านซ้าย ด้านละ $\alpha/2$ ตามลำดับ

ตัวอย่าง ต้องการทราบเวลาถัวเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของเวลาที่ใช้ส่งจดหมายจากเชียงใหม่ถึงกรุงเทพ และหา 95% ช่วงเชื่อมั่นของ σ โดยมีข้อมูลคือเวลาที่ใช้คิดเป็นชั่วโมงของจดหมาย 9 ฉบับ ดังนี้

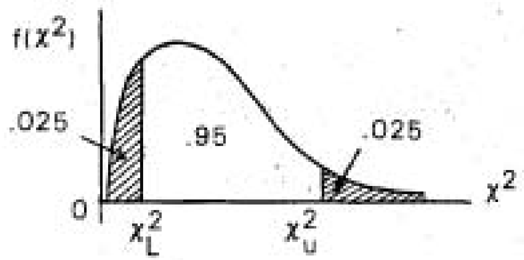
X = เวลา	$X - \bar{X}$	$(X - \bar{X})^2$
50	-9	81
45	-14	196
27	-32	1024
66	7	49
43	-16	256
96	37	1369
45	-14	196
90	31	961
69	10	100
ΣX = 531	0	4232 ← $\Sigma(X - \bar{X})^2$

$\bar{X} = \frac{531}{9} = 59$ ชั่วโมง $\qquad S^2 = \frac{\Sigma(X - \bar{X})^2}{n-1} = \frac{4232}{8} = 529$

$S = \sqrt{529} = 23$ ชั่วโมง

นั่นคือ จดหมายจากเชียงใหม่ถึงผู้รับในกรุงเทพใช้เวลาโดยเฉลี่ย 59 ชั่วโมง และมีส่วนเบี่ยงเบนมาตรฐาน 23 ชั่วโมง

ในการหา 95% ช่วงเชื่อมั่นของ σ ต้องหาของ σ^2 ก่อน คือ ต้องหา σ_L^2 และ σ_u^2 และต้องทราบค่าเปิดตาราง χ^2 ที่ df = 9 − 1 = 8 จากตารางจะได้ $\chi^2_{.025,\,8} = 17.525$ และ $\chi^2_{.975,\,7} = 2.180$ ในเมื่อ $\chi^2_{\alpha,\,\nu}$ คือค่าจากตาราง χ^2 โดยเปิดที่ df = ν และ α คือพื้นที่ด้านขวามือจนถึงปลายหางโดยนับจากจุด $\chi^2_{\alpha,\,\nu}$ โปรดดูรูปประกอบ $\chi_u^2 = \chi^2_{.025,\,8} = 17.525$ คือจากจุด 17.525 ถึงปลายหางขวามือมีพื้นที่ = .025 และ $\chi_L^2 = \chi^2_{.975,\,8} = 2.180$ คือจากจุด 2.180 ถึงปลายหางด้านขวามือ จะมีพื้นที่ = .975 จึงทำให้เหลือพื้นที่จากจุด 2.180 ถึงปลายหางด้านซ้ายมือ = .025 จึงทำให้เหลือพื้นที่ส่วนกลางโค้ง = .95



รูปที่ 10.4
แสดงการสร้างช่วงเชื่อมั่นของ σ^2

ดังนั้น

$$\sigma_L^2 = \frac{(n-1)S^2}{\chi_u^2} = \frac{8(529)}{17.535} = 241.35, \quad \sigma_L = \sqrt{241.35} = 15.54$$

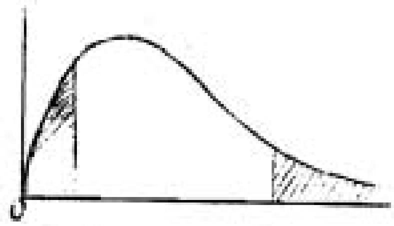
$$\sigma_u^2 = \frac{(n-1)S^2}{\chi_L^2} = \frac{8(529)}{2.180} = 1,941.28, \quad \sigma_u = \sqrt{1941.28} = 44.06$$

นั่นคือ 95% ช่วงเชื่อมั่นของ σ^2 คือ

$$241.35 < \sigma^2 < 1,941.28$$

และ 95% ช่วงเชื่อมั่นของ σ คือ

$$15.54 < \sigma < 44.06$$



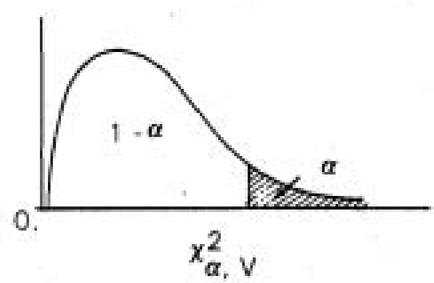
การทดสอบสมมติฐาน

เมื่อต้องการทดสอบสมมติฐานเกี่ยวกับความแปรปรวนของประชากร จะมีวิธีการดังนี้

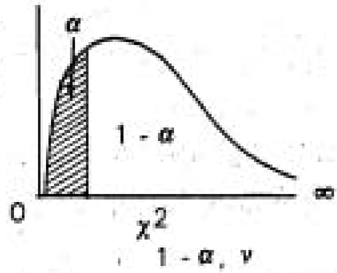
1. กำหนดสมมติฐานว่างเปล่า : Ho : $\sigma^2 = \sigma_0^2$
2. กำหนดสมมติฐานรอง : Ha : $\sigma^2 > \sigma_0^2$
   Ha : $\sigma^2 < \sigma_0^2$
   Ha : $\sigma^2 \neq \sigma_0^2$
3. กำหนดระดับนัยสำคัญ คือ α
4. เขตวิกฤตจะอยู่ภายใต้โค้ง χ^2 ที่มี $\nu = n - 1$ และขึ้นอยู่กับ Ha

4.1 เมื่อใช้ Ha : $\sigma^2 > \sigma_0^2$ เขตวิกฤตจะอยู่ปลายหางด้านขวามือ นั่นคือจะปฏิเสธ Ho เมื่อ $\chi^2 > \chi^2_{\alpha\ \nu}$

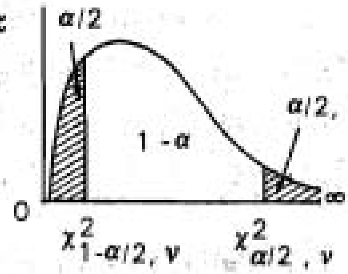
นั่นคือจะปฏิเสธเมื่อค่า χ^2 เป็นค่าที่โตเกินไปจนอยู่ในเขตวิกฤต



4.2 เมื่อใช้ Ha : $\sigma^2 < \sigma_0^2$ เขตวิกฤตจะอยู่ปลายทางด้านซ้ายมือ นั่นคือ จะปฏิเสธ Ho เมื่อ $X^2 < \chi^2_{1-\alpha,\nu}$ นั่นคือจะปฏิเสธ เมื่อค่า X^2 เป็นค่าที่เล็กเกินไปจนอยู่ในเขตวิกฤต



4.3 เมื่อใช้ Ha : $\sigma^2 \neq \sigma_0^2$ เขตวิกฤตจะมี 2 ด้าน ๆ ละ $\alpha/2$ จึงจะปฏิเสธ Ho เมื่อ $X^2 > \chi^2_{\alpha/2,\nu}$ หรือเมื่อ $X^2 < \chi^2_{1-\alpha/2,\nu}$ นั่นคือ เมื่อ X^2 เป็นค่าที่โตเกินไป หรือเล็กเกินไป



5. ตัวสถิติที่ใช้ทดสอบคือ

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

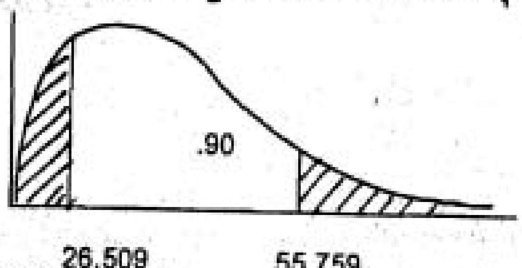
6. สรุปผล

ตัวอย่าง 1 ต้องการทดสอบคุณภาพของข้อสอบโดยต้องการให้มีส่วนเบี่ยงเบนมาตรฐาน 13 คะแนน คือ ไม่ต้องการให้ σ เป็นค่าเล็กเกินไป เพราะจะไม่สามารถแยกระหว่างนักเรียนเก่ง และไม่เก่งแสดงว่าข้อสอบง่ายเกินไปและไม่ต้องการให้ σ โตเกินไป เพราะแสดงว่า มีนักเรียนส่วนใหญ่ทำได้คะแนนต่ำ เพราะข้อสอบอาจยากเกินไป ค่าที่เหมาะคือ $\sigma = 13$ คะแนน เมื่อได้ลองใช้ข้อสอบกับนักเรียน 41 คน ได้คะแนนเฉลี่ย 72.7 และค่าเบี่ยงเบนมาตรฐาน 15.7 คะแนน ข้อสอบชุดนี้ดีหรือไม่?

1. Ho : $\sigma = 13$
2. Ha : $\sigma \neq 13$
3. $\alpha = .10$
4. จะปฏิเสธ Ho

เมื่อ $X^2 > \chi^2_{.05,40} = 55.759$

หรือ เมื่อ $X^2 < \chi^2_{.95,40} = 26.509$



5. $X^2 = \dfrac{(n-1)s^2}{\sigma_o^2} = \dfrac{40(15.7)^2}{(13)^2} = 58.34$

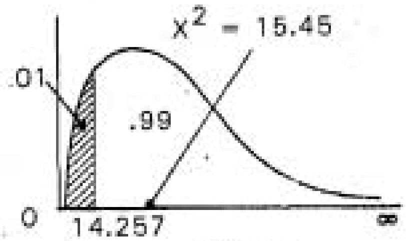
6. X^2 อยู่ในเขตวิกฤต จึงปฏิเสธ Ho แสดงว่า ส่วนเบี่ยงเบนมาตรฐานของข้อสอบชุดนี้โตเกินไป ข้อสอบคงยากเกินไป

ตัวอย่าง 2

โรงงานผลิตเครื่องชั่งน้ำหนักจะไม่ยอมปล่อยเครื่องชั่งที่มีความผันแปรสูงเกิน 1 ไมโครกรัม ออกจำหน่าย ขณะนี้ได้ผลิตเครื่องชั่งใหม่ 1 เครื่อง เจ้าหน้าที่ตรวจสอบคุณภาพได้ทดลองชั่งตุ้มเหล็กขนาด 500 กรัม ซ้ำกันในเวลาต่าง ๆ 30 ครั้ง ได้ค่าเบี่ยงเบนมาตรฐาน 0.73 ไมโครกรัม โรงงานควรปล่อยเครื่องชั่งนี้ออกสู่ตลาดหรือไม่ ถ้าใช้ $\alpha = .01$?

1. Ho : $\sigma^2 = 1$ **(สินค้าไม่ได้มาตรฐาน)**
2. Ha : $\sigma^2 < 1$ **(สินค้าได้มาตรฐาน)**
3. $\alpha = .01$
4. จะปฏิเสธ Ho เมื่อ $X^2 < X^2_{29,\ .99} = 14.257$
5. $X^2 = \dfrac{(n-1)s^2}{\sigma_o^2} = \dfrac{(30-1)(.73)^2}{(1)^2}$

   $= 15.45$



p-value > .01 จึงไม่ปฏิเสธ Ho

6. $X^2 = 15.45$ ไม่อยู่ในเขตวิกฤต จึงต้องยอมรับ Ho นั่นคือ ส่วนเบี่ยงเบนมาตรฐานมีค่า 1 กรัมขึ้นไป จึงควรส่งกลับคืนโรงงานเพื่อปรับปรุงคุณภาพ

แบบฝึกหัด

10.54 **กำหนดให้** $n = 16, \bar{x} = 32.5, S^2 = 16.9$ จงหา 95% ช่วงเชื่อมั่นของความแปรปรวนของประชากร $(9.22 < \sigma^2 < 40.48)$

10.55 กำหนดให้ค่าเบี่ยงเบนมาตรฐานของประชากรหนึ่ง = 310 ถ้าจากข้อมูลที่สุ่มมา 10 จำนวน ได้ค่าเบี่ยงเบนมาตรฐานเป็น 220 เราจะสรุปว่า ค่าเบี่ยงเบนมาตรฐานของประชากรต่างไปจาก 310 ด้วย $\alpha = .05$ ได้ไหม? ($X^2 = 4.53$, ยอมรับ Ho)

10.56 จากข้อมูลในอดีต พบว่า ความแปรปรวนของประชากรหนึ่ง = 48 ถ้าสุ่มข้อมูลมา 15 จำนวน ได้ความแปรปรวน 55 เราควรจะสรุปว่า ความแปรปรวนของประชากรเพิ่มสูงกว่าเดิม ด้วย α = .10 ไหม? (X^2 = 16.04, ยอมรับ Ho)

10.57 กำหนดให้ n = 12, S^2 = 224 จงสร้างช่วงเชื่อมั่น 90% ของความแปรปรวนประชากร

($112.41 < \sigma^2 < 645.70$)

10.58 ผู้จัดการฝ่ายผลิตเชื่อว่า พนักงานที่มีทักษะจะผลิตสินค้าได้มากกว่าพนักงานที่รับเข้าใหม่ แต่เชื่อว่า ความผันแปรของสินค้าที่ผลิตได้ของพนักงาน 2 กลุ่มนี้ไม่น่าแตกต่างกัน จากผลการศึกษาในอดีตพบว่า พนักงานเข้าใหม่ผลิตสินค้าได้โดยถัวเฉลี่ยชั่วโมงละ 20 หน่วย และมีความแปรปรวน 56 ถ้าพนักงานที่เข้ามา 5 ปีแล้ว (มีทักษะ) จำนวน 20 คน มีผลผลิตเฉลี่ยชั่วโมงละ 30 หน่วย และความแปรปรวน 28 หน่วย จะสรุปว่า พนักงาน 2 กลุ่ม มีความผันแปรในจำนวนผลิตต่างกันที่ α = .05 ไหม? (X^2 = 9.5, ยอมรับ Ho)

10.59 แบบสอบถามชนิดหนึ่งมีความแปรปรวนของประชากร 45 คะแนน ถ้าลองใช้แบบสอบถามนั้น กับพนักงานบริษัทหนึ่งจำนวน 24 คน ได้ความแปรปรวน 25 คะแนน ถ้าใช้ระดับนัยสำคัญ 5% จะแสดงว่า แบบสอบถามพนักงานของบริษัทนี้ มีความแปรปรวนต่ำกว่าค่าประชากรไหม? (X^2 = 12.78, ยอมรับ Ho)

10.60 ในการตรวจสอบไอเสียของรถ 25 คัน มีความแปรปรวน 54 จงหาช่วงเชื่อมั่น 95% ของความแปรปรวนที่แท้จริง ($32.92 < \sigma^2 < 104.51$)

10.61 ธนาคารแห่งหนึ่งต้องการลดค่าใช้จ่ายสำหรับการฝากเงินแบบออมทรัพย์ ซึ่งพบว่า ความแปรปรวนระหว่างการฝาก-ถอนเงินแต่ละครั้ง คือ 84 วัน ธนาคารต้องการลดความแปรปรวนเนื่องจากการฝากเงินในระยะสั้น จึงใช้นโยบายคิดค่าบริการจากรายการที่ถอนเกิน 1 ครั้ง ต่อ 1 เดือน และเมื่อสุ่มบัญชีเงินฝากออมทรัพย์มา 15 บัญชี พบว่า ความแปรปรวนเป็น 28 วัน ธนาคารจะสรุปว่า นโยบายใหม่ทำให้ความแปรปรวนลดลงด้วย α = .05 ได้ไหม?

(X^2 = 4.67, ปฏิเสธ Ho)

8. การอนุมานความแปรปรวนของ 2 ประชากร

(Inference about two population variances)

เมื่อเราต้องการเปรียบเทียบความแตกต่างของความแปรปรวนระหว่าง 2 ประชากร เราต้องนำความแปรปรวนจากตัวอย่างที่สุ่มมา 2 กลุ่ม เปรียบเทียบกัน แต่เราจะไม่เปรียบเทียบโดยใช้**ผลต่าง**เหมือนกับค่าเฉลี่ยและสัดส่วน เราจะใช้เทียบเป็น**อัตราส่วน** หรือ ratio แทน ดังตัวอย่างต่อไปนี้

ตัวอย่าง 1 นักเศรษฐศาสตร์เชื่อว่า รายได้ของผู้มีการศึกษาระดับวิทยาลัยจะมีความผันแปรสูงกว่ารายได้ของผู้ไม่จบวิทยาลัย เขาเก็บตัวอย่างคือผู้จบวิทยาลัย 21 คน พบว่า ค่าเบี่ยงเบนมาตรฐานของรายได้คือ $S_1 = 17{,}000$ บาท และจากกลุ่มตัวอย่างผู้ไม่จบวิทยาลัย 25 คน พบว่า มีค่าเบี่ยงเบนมาตรฐานของรายได้ S_2 เป็น 7,500 บาท จะสรุปว่าความเชื่อของเขาเป็นจริงที่ระดับนัยสำคัญ .01 ไหม?

ก่อนอื่นเราควรทราบขั้นตอนการทดสอบโดยทั่วไปก่อน ดังนี้

1. **ตั้งสมมติฐานว่างเปล่า:**

   $Ho : \sigma_1^2 = \sigma_2^2$ (หรือ $\sigma_1^2/\sigma_2^2 = 1$)

2. **ตั้งสมมติฐานรอง ซึ่งมี 3 อย่าง คือ**

   2.1 $Ha : \sigma_1^2 > \sigma_2^2$ (หรือ $\sigma_1^2/\sigma_2^2 > 1$)

   2.2 $Ha : \sigma_1^2 < \sigma_2^2$ (หรือ $\sigma_1^2/\sigma_2^2 < 1$)

   2.3 $Ha : \sigma_1^2 \neq \sigma_2^2$ (หรือ $\sigma_1^2/\sigma_2^2 \neq 1$)

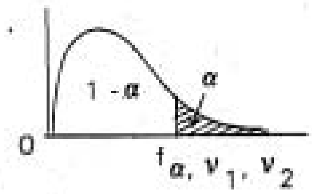
3. **กำหนดระดับนัยสำคัญ คือ** α

4. **แสดงเขตวิกฤต ซึ่งจะขึ้นอยู่กับสมมติฐานรอง เขตวิกฤตจะอยู่ภายใต้โค้งการแจกแจงแบบ -**

   f ที่มี $\nu_1 = n_1 - 1$ และ $\nu_2 = n_2 - 1$

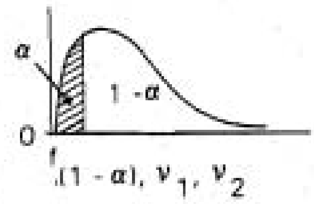
4.1 สำหรับ Ha : $\sigma_1^2 > \sigma_2^2$ เราจะปฏิเสธเมื่อค่าสถิติ F โตเกินไป คือเมื่อ

$F > f_{\alpha, \nu_1, \nu_2}$



4.2 สำหรับ Ha : $\sigma_1^2 < \sigma_2^2$ เราจะปฏิเสธ Ho เมื่อค่าสถิติ F เล็กเกินไป คือเมื่อ

$F < f_{(1-\alpha), \nu_1, \nu_2}$

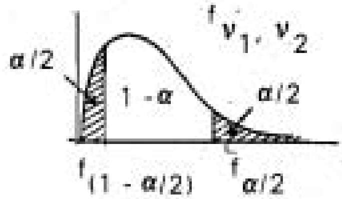


4.3 สำหรับ Ha : $\sigma_1^2 \neq \sigma_2^2$ เราจะปฏิเสธ Ho เมื่อค่าสถิติ F ที่คำนวณได้เป็นค่าที่โตเกินไป หรือ เล็กเกินไป นั่นคือ เมื่อ

$F > f_{\alpha/2, \nu_1, \nu_2}$

หรือ

$F < f_{1-\alpha/2, \nu_1, \nu_2}$



5. ค่าสถิติที่ใช้ทดสอบ คือ

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

แต่ σ_1^2/σ_2^2 หรือ $\sigma_2^2/\sigma_1^2 = 1$ ตามที่สมมุติไว้ใน Ho

จึงคำนวณ

$$F = \frac{S_1^2}{S_2^2}$$

สำหรับการสร้างช่วงเชื่อมั่นของ σ_1^2/σ_2^2 ต้องใช้ตัวสถิติ F ดังนี้

$$\boxed{F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \quad \text{มีการแจกแจงแบบค่าในตาราง } f_{\nu_1,\nu_2} \quad \nu_1 = n_1 - 1, \nu_2 = n_2 - 1}$$

$$= \left(\frac{S_1^2}{S_2^2}\right)\left(\frac{\sigma_2^2}{\sigma_1^2}\right)$$

$$\frac{1}{F} = \left(\frac{S_2^2}{S_1^2}\right)\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$$

ดังนั้น $\frac{\sigma_1^2}{\sigma_2^2} = \frac{1}{F} \Big/ \left[\frac{S_2^2}{S_1^2}\right] = \frac{1}{F}\,(S_1^2/S_2^2)$

ดังนั้น เราจึงหาช่วงเชื่อมั่นของ σ_1^2/σ_2^2 ดังนี้

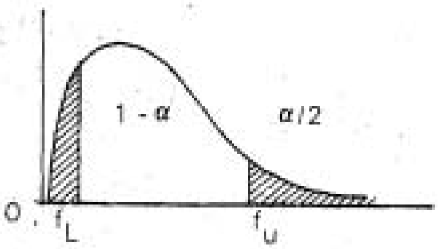
$(\sigma_1^2/\sigma_2^2)_L = \frac{(S_1^2/S_2^2)}{f_u}$ ขีดจำกัดล่าง

$(\sigma_1^2/\sigma_2^2)_u = \frac{(S_1^2/S_2^2)}{f_L}$ ขีดจำกัดบน

f_u ได้จากการเปิดตาราง F ที่ $\alpha = \alpha/2$

และ $\nu_1 = n_1 - 1, \nu_2 = n_2 - 1$

รูปที่ 10.5
แสดงการหาช่วงเชื่อมั่นของ σ_1^2/σ_2^2



ส่วนค่า f_L คือค่าที่ทำให้เหลือพื้นที่ทางขวามือ = 1 - α/2 จะไม่มีกำหนดให้ในตาราง เพราะตาราง f จะมีเฉพาะ α = .005, α = .01, α = .025, α = .05 และ α = .10 สมมุติเราต้องการสร้างช่วงเชื่อมั่น 90% ของ σ_1^2/σ_2^2 จุด f_L คือ $f_{1-.10/2} = f_{.95}$ จะไม่มีตาราง f ที่มี α โตถึง .95 เราจึงต้องใช้ความสัมพันธ์ ดังนี้

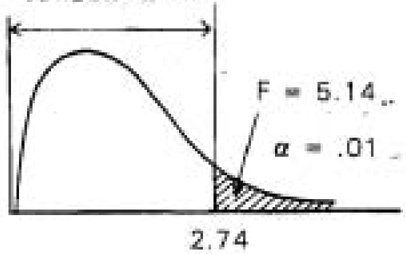
$$f_{(1-\alpha), \nu_1, \nu_2} = \frac{1}{f_{\alpha, \nu_2, \nu_1}}$$

$$\text{หรือ } f_{1-\alpha/2, \nu_1, \nu_2} = \frac{1}{f_{\alpha/2, \nu_2, \nu_1}}$$

จากตัวอย่างที่ 1

จบมหาวิทยาลัย	ไม่จบมหาวิทยาลัย
$n_1 = 21$	$n_2 = 25$
$S_1 = 17{,}000$	$S_2 = 7500$

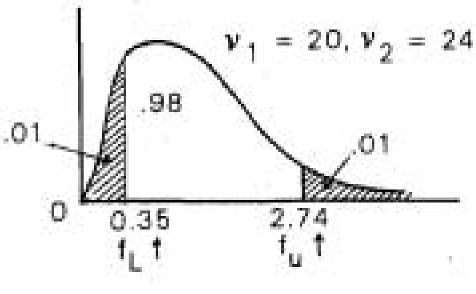
1. Ho : $\sigma_1^2 = \sigma_2^2$ **(หรือ $\sigma_1^2/\sigma_2^2 = 1$)**
2. Ha : $\sigma_1^2 > \sigma_2^2$ **(หรือ $\sigma_1^2/\sigma_2^2 > 1$)**
3. $\alpha = .01$
4. จะปฏิเสธ Ho เมื่อ $F > f_{.01;\, 20,\, 24} = 2.74$
5. $F = \dfrac{(17{,}000)^2}{(7{,}500)^2} = 5.14$



p-value < .01 จึงปฏิเสธ Ho

6. F = 5.14 อยู่ในเขตวิกฤต เราจึงปฏิเสธ Ho และยอมรับ Ha นั่นคือ ความเชื่อของนักเศรษฐศาสตร์ที่ว่ารายได้ผู้จบวิทยาลัย มีความผันแปรสูงกว่ารายได้ของผู้ไม่จบมหาวิทยาลัยเป็นความจริง

ถ้าเราต้องการสร้างช่วงเชื่อมั่น 98% ของ σ_1^2/σ_2^2 เราจะต้องหาค่า f_u และ f_L ก่อน และขอให้ดูรูปที่ 10.6



รูปที่ 10.6
แสดงการหาช่วงเชื่อมั่น 98%
ของ σ_1^2/σ_2^2 จากตัวอย่างที่ 1

$$f_u = f_{\alpha/2,\ \nu_1,\ \nu_2} = f_{.01;\ 20,\ 24} = 2.74$$

$$f_L = f_{(1-\alpha/2);\ \nu_1,\ \nu_2} = f_{.99;\ 20,\ 24}$$

$$= \frac{1}{f_{.01;\ 24,\ 20}}$$

$$= \frac{1}{2.86} = 0.35$$

ดังนั้น $$(\sigma_1^2/\sigma_2^2)_L = \frac{(S_1^2/S_2^2)}{f_u} = \frac{(17{,}000)^2/(7500)^2}{2.74}$$

$$= 1.88$$

และ $$(\sigma_1^2/\sigma_2^2)_u = \frac{(S_1^2/S_2^2)}{f_L} = \frac{(17{,}000)^2/(7{,}500)^2}{0.35}$$

$$= 14.68$$

นั่นคือ 98% ช่วงเชื่อมั่นของ σ_1^2/σ_2^2 คือ

$$1.88 < \sigma_1^2/\sigma_2^2 < 14.68$$

จะเห็นว่าถ้า $\sigma_1^2 = \sigma_2^2 = \sigma_1^2/\sigma_2^2 = 1$ เมื่อตรวจดู พบว่า 1 ไม่รวมอยู่ในช่วงเชื่อมั่น จึงสรุปว่า $\sigma_1^2 \neq \sigma_2^2$ ด้วยความเชื่อมั่น 98%

ตัวอย่างที่ 2 ต้องการเปรียบเทียบยาชา 2 ชนิด เพื่อใช้ในการถอนฟันโดยจะเปรียบเทียบความผันแปรของเวลาที่ใช้จนผู้ป่วยรู้สึกชาสมบูรณ์ ได้ข้อมูล ดังนี้

ยา A :	$n_1 = 31,$	$S_1^2 = 1296$
ยา B :	$n_2 = 41,$	$S_2^2 = 784$

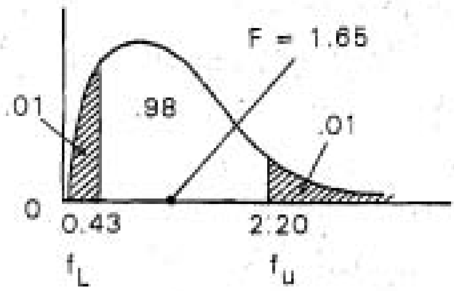
1. Ho : $\sigma_1^2 = \sigma_2^2$ (หรือ $\sigma_1^2/\sigma_2^2 = 1$)
2. Ha : $\sigma_1^2 \neq \sigma_2^2$ (หรือ $\sigma_1^2/\sigma_2^2 \neq 1$)
3. $\alpha = .02$
4. $\nu_1 = 30,\ \nu_2 = 40$

$$f_u = f_{.01;\, 30,\, 40} = 2.20$$

$$f_L = f_{.99;\, 30,\, 40} = \frac{1}{f_{.01,\, 40,\, 30}}$$

$$= \frac{1}{2.30}$$

$$= 0.43$$



นั่นคือจะปฏิเสธ Ho เมื่อ

$$F > f_u = 2.20$$

หรือ

$$F < f_L = 0.43$$

5. $F = \dfrac{S_1^2}{S_2^2} = \dfrac{1296}{784} = 1.65$

6. F = 1.65 อยู่ในเขตยอมรับ Ho จึงสรุปว่ายาทั้ง 2 ชนิด มีความผันแปรของเวลาที่ใช้จนผู้ป่วยรู้สึกช้าโดยสมบูรณ์ไม่ต่างกันอย่างมีนัยสำคัญนั่นคือ จะซื้อชนิดใดก็ให้คุณภาพเหมือนกัน ฉะนั้น ควรซื้อชนิดที่มีราคาถูก

และถ้าเราต้องการสร้างช่วงเชื่อมั่น 98% ของ σ_1/σ_2 จะมีวิธีดังนี้

$$(\sigma_1^2/\sigma_2^2)_L = \frac{(S_1^2/S_2^2)}{f_u} = \frac{1.65}{2.20} = 0.75$$

ดังนั้น

$$(\sigma_1/\sigma_2)_L = \sqrt{0.75} = 0.87 \longleftarrow \text{ขีดจำกัดล่าง}$$

และ $$(\sigma_1^2/\sigma_2^2)_u = \frac{(S_1^2/S_2^2)}{f_L} = \frac{1.65}{0.43} = 3.84$$

ดังนั้น $$(\sigma_1/\sigma_2)_u = \sqrt{3.84} = 1.96 \longleftarrow \text{ขีดจำกัดบน}$$

จึงได้ 98% ช่วงเชื่อมั่นของ σ_1/σ_2 ดังนี้

$$0.87 < \sigma_1/\sigma_2 < 1.96$$

จะเห็นว่าถ้า $\sigma_1 = \sigma_2$, $\sigma_1/\sigma_2 = 1$ เมื่อตรวจดูจะพบว่าอยู่ในช่วงเชื่อมั่น 98% จึงสรุปได้ว่า $\sigma_1 = \sigma_2$ ซึ่งสอดคล้องกับการทดสอบสมมติฐาน

แบบฝึกหัด

10.62 ถ้า $n_1 = 16, S_1 = 8.2$ และ $n_2 = 12, S_2 = 4.8$ ถ้าใช้ระดับนัยสำคัญ 5% จะสรุปว่าประชากรที่ 2 มีความแปรปรวนน้อยกว่าประชากรที่ 1 ได้ไหม? ($F = 2.92$, ปฏิเสธ Ho)

10.63 ถ้าความแปรปรวนของต้นทุนล้างอัดฟิล์ม A = 146 บาท โดยใช้ฟิล์มตัวอย่าง 15 ม้วน และความแปรปรวนของต้นทุนล้าง-อัดฟิล์ม B = 124 บาท จากตัวอย่าง 18 ม้วน เราจะสรุปด้วย $\alpha = .05$ ว่า ความแปรปรวนของฟิล์ม A สูงกว่าฟิล์ม B ได้ไหม? ($F = 1.18$, ยอมรับHo)

10.64 $n_1 = 12, S_1^2 = 1.96$ และ $n_2 = 10, S_2^2 = 3.64$

(ก) จงคำนวณค่า F เพื่อทดสอบความแตกต่างของความแปรปรวน ($F = 0.54$)

(ข) จงแสดงเขตวิกฤต เมื่อใช้ $\alpha = .10$

(ค) สรุปผลการทดสอบว่าอย่างไร

(ง) จงหาช่วงเชื่อมั่น 90% ของ σ_1^2/σ_2^2 ($0.17 < \sigma_1^2/\sigma_2^2 < 1.57$)

(จ) ผลสรุปจากการทดสอบในข้อ (ค) สอดคล้องกับช่วงเชื่อมั่นที่หาได้ได้ข้อ (ง) หรือไม่? จงอธิบาย

10.65 ในการเปรียบเทียบค่าเฉลี่ยจาก 2 ประชากรที่เป็นอิสระกัน และไม่ทราบค่า σ_1^2, σ_2^2 เรามักใช้การทดสอบแบบ t ซึ่งจะต้องมีข้อสมมุติว่า $\sigma_1^2 = \sigma_2^2$ ถ้านักทดลองผู้หนึ่งต้องการทดสอบอิทธิพลตอบสนองต่อตัวยาชนิดหนึ่ง เขาจึงแบ่งกลุ่มทดลองเป็น 2 กลุ่ม คือ กลุ่มควบคุม (control) กับกลุ่มที่ใช้ยา (treated) เขาได้ข้อมูล ดังนี้

กลุ่มควบคุม	กลุ่มใช้ยา
$n_1 = 32$	$n_1 = 35$
$S_1^2 = 18.6$	$S_2^2 = 27.8$

เขาอยากทราบว่า ถ้าจะมีความแตกต่างระหว่าง 2 กลุ่มนี้ ก็จะเป็นความแตกต่างเฉพาะค่าเฉลี่ยเท่านั้น แต่ประชากรทั้ง 2 มีความแปรปรวนไม่ต่างกัน ดังนั้น เขาจะใช้การทดสอบแบบ-

t ได้ไหม ถ้าใช้ $\alpha = .10$ ($F = S_1^2/S_2^2 = 0.67$, ยอมรับ Ho)

10.66 ผู้จัดการฝ่ายควบคุมคุณภาพสินค้าสงสัยว่า เครื่องแก้วที่โรงงานผลิต 2 ชนิด มีความผันแปรของความทนทาน (breaking points) ต่างกันเมื่อใช้เครื่องวัดความคงทนของแก้วคุณภาพดีและคุณภาพรอง ชนิดละ 25 ชิ้น พบว่า ความแปรปรวนของความทนทานแก้วคุณภาพดี = 5.2 และของแก้วคุณภาพรอง = 12.4 ถ้าใช้ $\alpha = .10$ จะสรุปว่าแก้ว 2 ชนิด มีความแปรปรวนแตกต่างกันอย่างมีนัยสำคัญไหม? ($F = 0.42$, ยอมรับ Ho)

10.67 ผู้จัดขายฝ่ายขาย 2 คน มีความเห็นไม่ตรงกันในข้อที่ว่า แม่บ้านในตัวเมืองมีความผันแปรของการจับจ่ายอาหารสูงกว่าแม่บ้านนอกตัวเมืองหรือไม่ เขาจึงสุ่มแม่บ้านทั้ง 2 ประเภทมาประเภทละ 65 คน พบว่า ความแปรปรวนของจำนวนวันที่ใช้จับจ่ายซื้ออาหารของแม่บ้านในกรุง = 9.6 และขอแม่บ้านนอกกรุง = 4.2 จงสรุปผลโดยใช้ $\alpha = .10$ ($F = 2.29$, ปฏิเสธ Ho)

แบบฝึกหัดทบทวน

10.68 ผู้ผลิตโทรทัศน์ต้องการทราบความผันแปรของราคาขายปลีกโทรทัศน์ชนิดขาว-ดำ ขนาด 19 นิ้ว เมื่อสุ่มร้านค้าตัวอย่างมา 20 แห่ง พบว่า มีราคาขายเฉลี่ยเครื่องละ 1,900 บาท และค่าเบี่ยงเบนมาตรฐาน 160 บาท จงหาช่วงเชื่อมั่น 90% ของความแปรปรวนของประชากร ($16,135.88 < \sigma^2 < 48,077.49$)

10.69 จงหาข้อสรุปที่เหมาะสมสำหรับข้อมูลในตารางคอนทินเจ็นซีต่อไปนี้ โดยใช้ $\alpha = .05$
(คำตอบ $X^2 = 8.16$, ยอมรับ Ho)

ทัศนคติต่อกฎหมายทำแท้งเสรี

อาชีพ	เห็นด้วย	ไม่ออกความเห็น	ไม่เห็นด้วย
ผู้ใช้แรงงาน	18	12	36
ผู้บริหาร	11	15	42
นักวิชาชีพ	24	8	32

10.70 ธนาคารแห่งหนึ่งต้องการทราบว่า ระบบการใช้เครื่องจักรอัตโนมัติโดยสมบูรณ์ โดยไม่ต้องสอบถามพนักงานเลย จะเป็นที่ยอมรับในกลุ่มผู้มีรายได้ต่าง ๆ เพียงไร จึงได้ทดลองใช้ระบบอัตโนมัติสมบูรณ์ในสาขาต่าง ๆ 3 สาขา และได้จำแนกลูกค้าตามระดับรายได้เป็น 3 กลุ่ม ได้ผลการสำรวจ ดังนี้

จำนวนการตอบรับของผู้มีรายได้

ทัศนคติ	ต่ำ	ปานกลาง	สูง	รวม
ชอบ	30	45	23	98
ไม่ชอบ	30	35	27	92

จงสรุปผลโดยใช้ระดับนัยสำคัญ 0.05

(คำตอบ : X^2 = 1.38, ยอมรับ Ho)

10.71 เราจะใช้การแจกแจงความน่าจะเป็นแบบใดสำหรับการทดสอบต่อไปนี้

ก) เปรียบเทียบสัดส่วนของ 2 ประชากร

ข) เปรียบเทียบความแปรปรวนของ 1 ประชากร

ค) เปรียบเทียบค่าเฉลี่ยระหว่าง 3 ประชากรขึ้นไป

ง) เปรียบเทียบค่าเฉลี่ยของ 2 ประชากร จากตัวอย่างที่ไม่เป็นอิสระกัน

10.72 เราจะใช้การแจกแจงความน่าจะเป็นแบบใด สำหรับการทดสอบต่อไปนี้

ก) เปรียบเทียบค่าเฉลี่ยของตัวอย่างขนาดเล็ก 2 กลุ่ม ซึ่งมาจากประชากรซึ่งไม่ทราบค่าความแปรปรวน

ข) เปรียบเทียบความแปรปรวนของ 2 ประชากร

ค) ค่าเฉลี่ยของประชากร 1 ประชากร โดยใช้ข้อมูลจากตัวอย่างขนาดใหญ่

ง) เปรียบเทียบสัดส่วนของ 3 ประชากรขึ้นไป

10.73 ผู้จัดการฝ่ายผลิตทดลองผลิตสินค้าด้วยวิธีการผลิต 3 วิธี เพื่อต้องการเปรียบเทียบต้นทุนการผลิต ได้ข้อมูลดังนี้

ต้นทุนการผลิตต่อหน่วย

วิธีที่ 1	6.50	7.20	6.80	6.90	6.40	7.30
วิธีที่ 2	4.90	5.30	4.80	4.60	5.90	5.00
วิธีที่ 3	6.10	5.90	5.80	6.10	6.00	5.70

จงใช้ระดับนัยสำคัญ .01 ตรวจสอบว่าต้นทุนการผลิตต่อหน่วยของวิธีต่าง ๆ มีความแตกต่างอย่างมีนัยสำคัญหรือไม่? (F = 37.74, ปฏิเสธ Ho)

10.74 บริษัทผลิตแผ่นป้ายโฆษณาต้องการทราบว่า มีความแตกต่างในขนาดของจราจรซึ่งผ่านจุดที่จะตั้งแผ่นป้ายโฆษณา 3 จุด หรือไม่ เพราะบริษัทจะคิดค่าบริการตามขนาดของจราจรที่ผ่านจุดโฆษณา บริษัทจึงสำรวจขนาดของจราจรโดยเลือกเวลาต่าง ๆ แบบสุ่ม และนับจำนวนรถที่ผ่านไป-มาในช่วง 5 นาที ได้ข้อมูลดังนี้

ขนาดของจราจร

จุดที่ 1	30	45	26	44	18	38	42	29	
จุดที่ 2	24	33	31	16	31	13	12	25	27
จุดที่ 3	35	47	43	46	27	31	21		

ขนาดของการจราจรที่ผ่าน 3 จุดเหมือนกันไหม? $\alpha = .05$ (F = 4.26, ปฏิเสธ Ho)

10.75 บริษัทโฆษณาอีกแห่งหนึ่งกำลังพิจารณาเลือกซื้อเวลาสำหรับโฆษณาทางโทรทัศน์ระหว่างรายการโทรทัศน์ 3 โปรแกรม จากการสุ่มมา 6 สัปดาห์ เพื่อดูว่าแต่ละโปรแกรม มีสัดส่วนของผู้ชมที่ตกอยู่ใน "ตลาดเป้าหมาย" กี่เปอร์เซนต์ ได้ข้อมูลดังนี้

เปอร์เซนต์

โปรแกรม 1	85	71	78	89	74	95
โปรแกรม 2	65	77	84	75	71	96
โปรแกรม 3	76	86	77	76	84	85

สัดส่วนผู้ชมที่ตกอยู่ใน "ตลาดเป้าหมาย" ของแต่ละโปรแกรมแตกต่างกันหรือไม่ เมื่อใช้ $\alpha = .05$ (F = 0.33, ยอมรับ Ho)

10.76 ท่านจะสรุปผลจากตารางคอนทินเจนซีข้างล่างนี้ว่าอย่างไร เมื่อใช้ $\alpha = .05$?

ระดับรายได้

การฟังเทศน์	ต่ำ	ปานกลาง	สูง
ไม่เคย	28	52	16
บางโอกาส	25	66	14
เป็นประจำ	18	73	8

(X^2 = 8.33, ยอมรับ Ho)

10.77 ท่านจะสรุปผลว่าอย่างไรสำหรับข้อมูลในตารางข้างล่าง โดยใช้ α = .01

กลุ่มอายุ

ประเภทของรถที่ขับขี่	16-21	22-30	31-45	46+
รถสปอร์ท	10	15	12	8
รถขนาดเล็ก	5	7	6	8
รถขนาดกลาง	12	14	20	25
รถขนาดใหญ่	8	12	21	25

(X^2 = 9.42, ยอมรับ Ho)

10.78 จำนวนเรือบินที่แวะลง ณ ท่าอากาศยานแห่งหนึ่งในช่วงเวลา 30 นาที วันเวลาที่สุ่มมา 250 คาบเวลา (30 นาที) ดังนี้

จำนวนเรือบิน (ต่อ 30 นาที)	0	1	2	3	4 ขึ้นไป
จำนวนคาบเวลา	47	56	71	44	32

ให้ทดสอบสอบว่าข้อมูลมีการแจกจงแบบปัวซองที่มี μ = 2 หรือไม่เมื่อใช้ α = .05?
(X^2 = 7.72, ยอมรับ Ho)

10.79 มีหลักฐานทางสังคมวิทยาที่แสดงว่าทัศนคติจากกลุ่มหญิงจะมีความผันแปรสูงกว่าทัศนคติจากกลุ่มชาย จากการสำรวจทัศนคติโดยสำนักงานวิจัยขนาดใหญ่ พบว่า ชายมีค่าเบี่ยงเบนมาตรฐานของทัศนคติ 15 คะแนน ถ้านักสังคมวิทยาผู้หนึ่งลองสำรวจทัศนคติจากหญิง 30 คน พบว่า มีความแปรปรวน 360 คะแนน ถ้าใช้ α = .05 จะมีเหตุผลเพียงพอที่จะเชื่อว่าทัศนคติของหญิงมีความความผันแปรสูงกว่าชายไหม? (X^2 = 46.4, ปฏิเสธ Ho)

10.80 นักจิตวิทยาสังคมได้สัมภาษณ์ 150 คน เพื่อวัดทัศนคติต่อ "สิทธิสตรี" โดยใช้คำตอบเพียง 2 อย่าง คือ เห็นด้วย กับไม่เห็นด้วย ข้อมูลข้างล่างคือจำนวนคำตอบที่ "เห็นด้วย" จำแนกตามกลุ่มอายุ และเธอใช้ค่าเฉลี่ยและความแปรปรวนจากตัวอย่างเป็นค่าประมาณของ-μ และ σ^2 ทำให้เธอทราบจำนวนคาดหมายถ้าข้อมูลมีการแจกแจงแบบปกติ ข้อมูลที่ได้มามีการแจกแจงแบบปกติหรือไม่ ถ้าใช้ α = .025? (X^2 = 11.75, ปฏิเสธ Ho)

จำนวนรายการที่เห็นด้วย

	10 หรือต่ำกว่า	11-12	13-14	15-16	17-18	19+
จำนวนบุคคลในแต่ละกลุ่ม	8	27	53	32	26	4
จำนวนบุคคลจากการแจกแจงแบบปกติ	14	26	41	36	22	11

10.81 นักจิตวิทยาเชื่อว่า ความเครียดและความกังวลมีอิทธิพลต่อผลการสอบของบุคคล เขาจึงแบ่งผู้ทดลองเป็น 2 กลุ่ม ๆ ละ 18 คน ให้กลุ่มหนึ่งทำแบบทดสอบในภาวะไม่ตึงเครียด และอีกกลุ่มทำการสอบข้อสอบชุดเดียวกับกลุ่มแรกแต่ให้มีภาวะตึงเครียด และผู้ทดลองค่อนข้างมั่นใจว่าภาวะตึงเครียดจะมีอิทธิพลในการเพิ่มความผันแปรของคะแนนสอบ เพราะเชื่อว่า นักเรียนบางคนสามารถทำสอบได้ดีในภาวะตึงเครียดมากกว่าภาวะปกติ ในขณะที่นักเรียนอีกหลายคนทำสอบไม่ได้ดีนักในภาวะตึงเครียด ถ้าความแปรปรวนของกลุ่มไม่ตึงเครียดคือ $S_1^2 = 22.8$ และของกลุ่มตึงเครียด คือ $S_2^2 = 78.5$ ข้อมูลนี้ สนับสนุนความเชื่อของนักจิตวิทยาไหม เมื่อใช้ $\alpha = .05$? (F = 3.44, ปฏิเสธ Ho)

10.82 ในการพัฒนายากล่อมประสาท จะต้องตรวจสอบอิทธิพลของยาต่อการใช้เครื่องจักร และการขับขี่รถ โรงงานผลิตยาได้ทดลองยาดังกล่าว 4 ชนิด โดยศึกษาผลกระทบต่อการขับขี่รถ โดยให้ผู้เข้ารับการทดลองขับขี่ในบริเวณทดลอง คะแนนที่ได้จะแสดงจำนวนความผิดพลาดในระหว่างขับขี่รถ ถ้าผิดมากจะมีคะแนนสูง ได้คะแนนดังนี้

ยา 1	230	258	239	241	
ยา 2	285	276	263	274	
ยา 3	215	232	204	247	226
ยา 4	241	253	237	246	210

จงใช้ $\alpha = .05$ ทดสอบอิทธิพลของยาทั้ง 4 ชนิดต่อการขับขี่

(F = 9.6, ปฏิเสธ Ho)