บทที่ 9
การวิเคราะห์ความแปรปรวน

ในบทก่อนๆเราได้เรียนรู้วิธีที่จะใช้การแจกแจง t ในการทดสอบสมมุติฐานระหว่างค่าเฉลี่ยของสองประชากรที่ไม่แตกต่างกัน สมมุติว่า เราต้องการรู้เกี่ยวกับผลสัมพัทธ์ของการปฏิบัติการ (treatment) ของสามกลุ่มหรือมากกว่า ศึกษาความแตกต่างประสิทธิผลของการรักษาสามวิธี ว่าแตกต่างกันหรือไม่ หรือศึกษาผลของการให้การเสริมแรงสี่วิธีระหว่างเด็กประถมศึกษา หรือศึกษาผลการทดสอบยาสำหรับใช้กับคนไข้ที่มีผลต่อความสามารถในการจำ

ในแต่ละตัวอย่างเหล่านี้ เราสามารถใช้การทดสอบ t เพื่อเปรียบเทียบระหว่างแต่ละคู่ของค่าเฉลี่ย อย่างไรก็ตาม วิธีนี้ไม่เหมาะสมที่จะใช้เป็นวิธีทั่วไป พิจารณาตัวอย่างที่เกี่ยวข้องกับการทดลองยากับคนไข้เจ็ดวิธี ขั้นแรกเราจะพบว่ามี 21 การทดสอบ t ถ้ายาแต่ละตัวเปรียบเทียบยาอื่นๆแต่ละตัว ขั้นที่สอง กับการทดสอบหลายครั้ง จะเกิดความคลาดเคลื่อนเพิ่มขึ้นอย่างน้อยก็ความคลาดเคลื่อนแบบที่ I นั่นคือจะได้รับ “แตกต่างอย่างมีนัยสำคัญ” เมื่อความแตกต่างนั้นไม่จริง สมมุติว่าไม่มียาตัวใดมีผล ถ้าเรากำหนดระดับนัยสำคัญที่ .05 สำหรับการทดสอบแต่ละกรณี ความน่าจะเป็นของความคลาดเคลื่อนแบบที่1 คือ 0.05 สำหรับแต่ละกรณี อย่างไรก็ตามเมื่อพิจารณาเป็นกลุ่ม ความน่าจะเป็นที่อย่างน้อยหนึ่งกรณีจากหลายๆกรณีจะพิสูจน์ว่าล้มเหลว ที่เป็นบวกมากกว่า 0.05 และเพิ่มขึ้น ขึ้นกับการทดสอบที่เพิ่มขึ้น อันที่จริง ถ้าทั้ง 21 การทดสอบเป็นอิสระต่อกัน ความน่าจะเป็นของอย่างน้อยหนึ่งกรณีที่จะเกิดความล้มเหลวเป็น 0.64 [คำนวณโดย $p = 1 - (.95)^{21}$] เมื่อการทดสอบไม่เป็นอิสระ ดังแสดงในตัวอย่าง มันเป็นไปไม่ได้ที่จะคำนวณความน่าจะเป็นที่แน่นอนของการกระทำที่อาจก่อให้เกิดความคลาดเคลื่อนแบบที่I ค่อนข้างสูง ดังนั้นถ้าเราจึงควดหาผลลัพธ์อันหนึ่งที่มีนัยสำคัญที่ระดับ 0.05 ระหว่างการทดสอบ 21 ครั้ง

ขั้นที่สามในการเปรียบเทียบใดๆ ใช้เพียงข้อมูลที่จัดหาโดยเปรียบเทียบกลุ่ม เมื่อการทดสอบทั้ง21 ครั้งสมบูรณ์ มีข้อมูล 21 ชุด ซึ่งจะมีความแตกต่างระหว่างผลของการใช้ยา 7 ชนิด

Ser Ronald Fisher พัฒนาคำตอบที่ดีกว่าสำหรับปัญหาเช่นนี้ เทคนิคนี้เรียก การวิเคราะห์ความแปรปรวน (analysis of variance) เขียนแทนด้วย ANOVA เทคนิคนี้ช่วยให้เราเปรียบเทียบหลายๆค่าเฉลี่ยในเวลาเดียวกัน กับระดับของนัยสำคัญที่จำเพาะเจาะจงโดยผู้สืบสวนสอบสวน การวิเคราะห์ความแปรปรวนแท้จริงเป็นเทคนิคการออกแบบเพื่อช่วยการทดสอบสมมุติฐาน ในบทนี้เราจะพัฒนารายละเอียดของรูปแบบที่ง่ายที่สุด ที่เรียกการวิเคราะห์ความแปรปรวนทางเดียว (one-way analysis of variance)

1 การทดสอบสมมุติฐาน

ในการทดสอบ t มีสองเงื่อนไขการดำเนินการ(treatment) และเราเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มทั้งสอง การวิเคราะห์ความแปรปรวนทางเดียวอนุญาตให้เราเปรียบเทียบค่าเฉลี่ยของสองกลุ่มหรือมากกว่าในเวลาเดียวกัน มันมีความสัมพันธ์ใกล้ชิดกับการทดสอบ t สำหรับกรณีของสองกลุ่ม ซึ่งนำไปสู่ข้อสรุปที่แน่นอนเหมือนกันกับการทดสอบ t การทดสอบ t อาจถูกคิดเสมือนว่าเป็นกรณีพิเศษของการวิเคราะห์ความแปรปรวนทางเดียว ในทางกลับกัน การวิเคราะห์ความแปรปรวนทางเดียวอาจพิจารณาว่าเป็นการขยายของการทดสอบ t สำหรับปัญหาที่เกี่ยวข้องมากกว่าสองกลุ่ม

ในการวิเคราะห์ความแปรปรวนทางเดียว อาจมีเงื่อนไขการปฏิบัติของสองกลุ่มหรือมากกว่า บ่อยครั้งจะอ้างอิงถึงระดับความแตกต่างของตัวแปรอิสระ เราจะใช้ k แทนจำนวนของเงื่อนไขการปฏิบัติ การปฏิบัติ k กลุ่มอาจแยกแยะด้วยตัวอักษร เช่น A,B,C,... และค่าเฉลี่ยของประชากรอาจใช้ μ_A, μ_B, μ_C,... ถ้าการปฏิบัติที่แตกต่างกัน ไม่มีผลแตกต่างบนตัวแปรภายใต้การสังเกต แล้วเราคาดหวังว่าค่าเฉลี่ยของประชากรเหล่านี้จะเท่ากัน การสืบสวนว่าการเบี่ยงเบนในเงื่อนไขการปฏิบัติทำให้เกิดความแตกต่าง เราจะทดสอบสมมุติฐาน

$$H_0: \mu_A = \mu_B = \mu_C = \mu_K$$

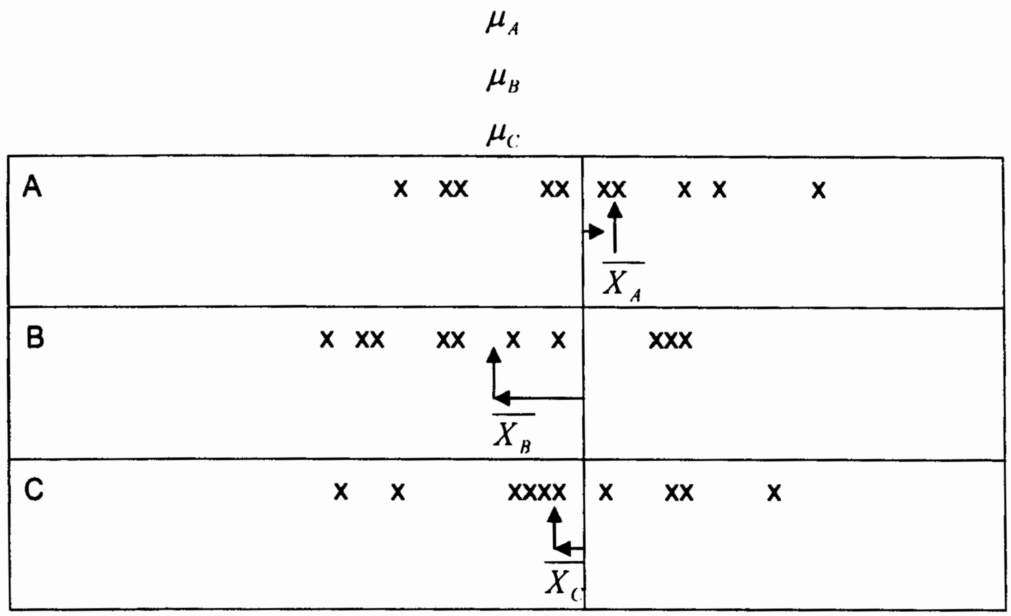
สมมุติฐานว่างใน ANOVA บ่อยครั้งอ้างอิงถึงการตั้งสมมุติฐานหลายอย่าง (ยกตัวอย่าง ครอบคลุมหลายสถานการณ์ในแต่ละครั้ง) และตัวANOVA เองเปรียบเสมือนการทดสอบหลายสถานการณ์ สมมุติฐานแย้ง บ่อยครั้งกำหนดง่ายๆ เช่น "ไม่ H_0" คือ ค่าเฉลี่ยประชากรแตกต่างกันในบางวิธี นั่นคือ อย่างน้อยหนึ่งกลุ่มของตัวอย่างที่ต่างไปจากประชากรด้วยค่าเฉลี่ย (μ) แตกต่างจากกลุ่มอื่นๆ

ในการทดสอบสมมุติฐานของ ไม่มีความแตกต่างระหว่างสองค่าเฉลี่ย ความแตกต่างจะถูกทำระหว่างสมมุติฐานแย้งแบบมีทิศทางหรือไม่มีทิศทาง เมื่อจำนวนค่าเฉลี่ยมากกว่าสองกลุ่ม ในการวิเคราะห์ความแปรปรวนของหลายกลุ่ม H_0 อาจล้มเหลวในกรณีใดก็ได้ ยกตัวอย่าง ค่าเฉลี่ยของสองกลุ่มหรือมากกว่าสองกลุ่ม อาจคล้ายกันหรือแตกต่างกันบางคู่หรือแตกต่างกันทั้งหมด

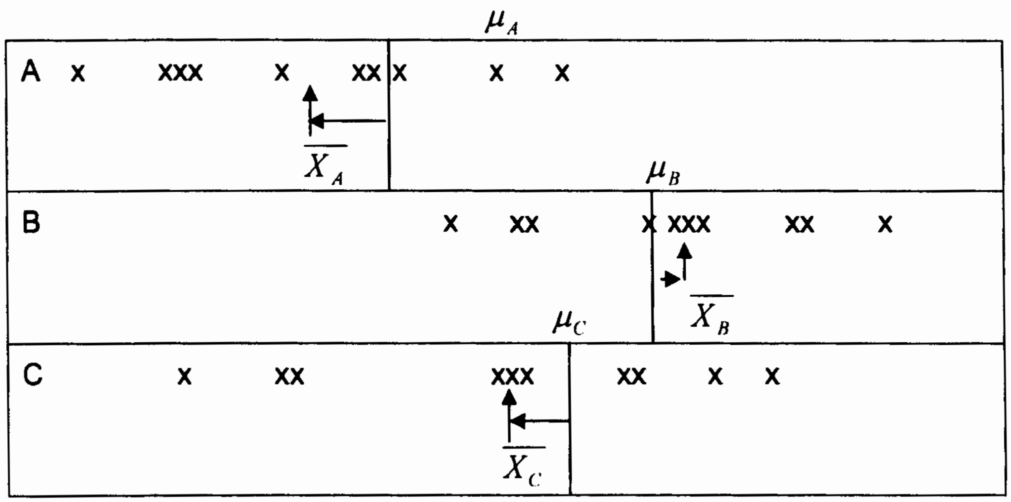
ตรรกะของการวิเคราะห์ความแปรปรวนทางเดียว : การเบี่ยงเบนภายในกลุ่มและระหว่างกลุ่ม

เราจะเริ่มพัฒนาการวิเคราะห์ความแปรปรวนทางเดียวของกลุ่มต่างๆ ANOVA ของการวัดซ้ำซึ่งจะแนะนำต่อไป

สมมุติว่าเราเลือกกลุ่มตัวอย่างขนาด 10 สามกลุ่มในแต่ละการสุ่ม และกำหนดการปฏิบัติแตกต่างกันสามแบบ ถ้าไม่มีผลการปฏิบัติที่แตกต่างกัน สมมุติฐานว่างคือ $\mu_A = \mu_B = \mu_C$ เป็นจริง และการแจกแจงของคะแนนในสามกลุ่มตัวอย่างอาจปรากฏดังแสดงในรูป 9.1 ถ้ามีผลการปฏิบัติแตกต่างกัน H_0 เป็นเท็จ และสามกลุ่มตัวอย่างอาจปรากฏ ดังรูป 9.2 เมื่อ μ_A, μ_B, μ_C เป็นค่าที่แตกต่างกัน เมื่อเปรียบเทียบรูป 9.1 และ 9.2 ซึ่งพิจารณากับสองแบบของการเบี่ยงเบน (เราใช้คำว่าเบี่ยงเบนในที่นี้ในความรู้สึกถึงค่าเฉลี่ยเพียงว่าค่าเปลี่ยนแปลง)



รูป 9.1



รูป 9.2

1. การเบี่ยงเบนภายในกลุ่ม ภายในแต่ละกลุ่มตัวอย่าง คะแนนของแต่ละสิ่งเบี่ยงเบนรอบค่าเฉลี่ยกลุ่มตัวอย่าง เราจะเรียกว่าการเบี่ยงเบนภายในกลุ่ม(within-groups variation) การเบี่ยงเบนตามที่คุณได้เรียนรู้เกี่ยวกับอิทธิพลของการเบี่ยงเบนการสุ่ม ซึ่งเป็นการสะท้อนโดยตรงของ การเบี่ยงเบนประจำตัว (inherent variation) ระหว่างแต่ละ

คะแนนที่ได้รับการปฏิบัติอย่างเดียวกัน เราสามารถนำเสนอสถานการณ์เดียวกันกับทุก ๆ สิ่งในกลุ่มและยังคงสังเกตความเบี่ยงเบนของปฏิกิริยา (reaction times) หรือเราสามารถใช้ร่องรอย และยังคงสังเกตความแตกต่างของแต่ละสิ่งในอัตราการเรียนรู้ มันเป็นข้อเท็จจริงของชีวิตที่ภายใต้เงื่อนไขอย่างเดียวกัน แต่ละบุคคลจะแปรเปลี่ยนในการกระทำ

2. การเบี่ยงเบนระหว่างกลุ่ม ค่าเฉลี่ยของกลุ่มตัวอย่างแปรเปลี่ยนระหว่างกลุ่มเหล่านั้น เราไม่คาดหวังว่า $\bar{x}_A, \bar{x}_B$ และ $\bar{x}_C$ ทั้งหมดมีค่าเหมือนกัน แม้ว่าเมื่อ $\mu_A = \mu_B = \mu_C$ สิ่งนี้ตามที่เราได้เรียนเกี่ยวกับการชักตัวอย่างแบบสุ่ม มันเป็นสิ่งสำคัญที่จะสำนึกว่าการเบี่ยงเบนระหว่างกลุ่มเป็นการสะท้อนของธรรมชาติการเบี่ยงเบนระหว่างแต่ละบุคคล พิจารณาว่าอะไรจะเกิดขึ้นถ้า $\mu_A = \mu_B = \mu_C$ และไม่มีธรรมชาติการเบี่ยงเบน ทุกคนในสามกลุ่มประชากรจะได้คะแนนเหมือนกัน ดังนั้นค่าเฉลี่ยของสามกลุ่มตัวอย่างจะไม่เบี่ยงเบนจากแต่ละคะแนน อีกด้านหนึ่ง ยิ่งมีธรรมชาติการเบี่ยงเบนมากระหว่างคะแนนของแต่ละคะแนน ก็ยิ่งมีโอกาสได้ค่าเฉลี่ยของกลุ่มตัวอย่างที่เบี่ยงเบนจากอีกค่าเฉลี่ยหนึ่ง

2. ตรรกะของการวิเคราะห์ความแปรปรวน

ในตอนนี้คุณอาจถูกถามว่าคะแนนเบี่ยงเบนอย่างไรภายในกลุ่มและระหว่างกลุ่ม ในขณะที่ทดสอบสมมุติฐานเกี่ยวกับค่าเฉลี่ย คำตอบคือว่าสองชนิดของการเบี่ยงเบนนี้ก่อให้เกิดมูลฐานสำหรับการทดสอบ $H_0 : \mu_A = \mu_B = \mu_C$

ก่อนอื่นเราจะพิจารณาการเบี่ยงเบนภายในกลุ่มของคะแนนแต่ละตัวเกี่ยวกับค่าเฉลี่ยของตัวอย่าง ถ้าเราสมมุติว่าความเป็นเอกพันธ์ของความแปรปรวนและคำนวณ s^2 สำหรับแต่ละตัวอย่าง แต่ละผลลัพธ์เป็นการประมาณของความแปรปรวนประชากร เพื่อให้ได้รับการประมาณที่ดี เราสามารถรวม (pool) หลาย ๆ(k) การประมาณตัวอย่างและได้รับค่าเฉลี่ย(สมมุติว่า $n_A = n_B = n_C = n_k$)

$$\frac{\sum s^2}{k} \xrightarrow{\text{ประมาณ}} \sigma^2$$

ในตอนนี้เรามีหนึ่งค่าประมาณของความแปรปรวนประชากร สังเกตว่ามันไม่ขึ้นกับ H_0 ว่าเป็นจริงหรือไม่

ในตอนนี้ให้เราพิจารณาการเบี่ยงเบนระหว่างกลุ่มของค่าเฉลี่ย สมมุติว่า H_0 เป็นจริง ถ้ามันเป็น k กลุ่มมาจากประชากรที่เป็นเอกลักษณ์กัน ซึ่งสมมูลกับที่จะกล่าวว่ามันมาจากประชากรเดียวกัน ระลึกว่าการแจกแจงการชักตัวอย่างของ $\overline{X}$ ความแปรปรวนของ $\overline{X}$ เท่ากับความแปรปรวนของประชากรที่แบ่งโดยขนาดของตัวอย่าง

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

การประมาณที่ดีที่สุดของ $\sigma^2_{\overline{X}}$ คือ $s^2_{\overline{X}}$ จากนี้เราสามารถได้รับการประมาณอื่นของ σ^2 ถ้า H_0 เป็นจริง

$$ns^2_{\overline{X}} \xrightarrow{\text{ประมาณ}} \sigma^2$$

สิ่งนี้ทำให้ได้สองการประมาณของความแปรปรวนของประชากร ถ้าสมมุติฐานว่างเป็นจริง ทั้งสองกลุ่มสะท้อนเพียงการเบี่ยงเบนซึ่งมีอยู่เดิม อย่างไรก็ตาม ถ้า H_0 เป็นเท็จโดยวิธีใดวิธีหนึ่ง(เช่น มี $\overline{X}$ อย่างน้อยหนึ่งตัวที่ตกลงไปจากความแตกต่างของประชากร) การประมาณค่าระหว่างกลุ่ม จะใหญ่ขึ้นเพราะว่ามันจะส่งผลต่อความแตกต่างของผลการปฏิบัติ ถ้าการประมาณระหว่างกลุ่มสูงกว่าภายในกลุ่ม ซึ่งจะก่อให้เกิดความไม่สมเหตุสมผลสำหรับมัน เราจะปฏิเสธ H_0

3 การแบ่งกั้นของผลบวกของกำลังสอง

เมื่อเปรียบเทียบสองกลุ่มตัวอย่างหรือมากกว่า ค่าเฉลี่ยของทุกคะแนนเรียกว่าค่าเฉลี่ยรวม (the grand mean) ใช้สัญลักษณ์ $\overline{\overline{X}}$ เราจะได้การบวกของทุกคะแนนในทุกกลุ่มและหารด้วยผลรวมจำนวนคะแนน คะแนนใดๆ ในกรุปใดสามารถแสดงด้วยความเบี่ยงเบนจากค่าเฉลี่ยรวม, $\left(X - \overline{\overline{X}}\right)$ คะแนนเบี่ยงเบนไปจากค่าเฉลี่ยรวมเท่าไร ถ้าเราแบ่งเป็นสองส่วน (1) คะแนนเบี่ยงเบนจากค่าเฉลี่ยของกลุ่มตัวอย่างของมันเท่าไรนั่นคือหา$\left(X - \overline{X}\right)$ และ (2) คะแนนของค่าเฉลี่ยของกลุ่มตัวอย่างเบี่ยงเบนไปจากค่าเฉลี่ยรวมเท่าไรนั่นคือหา $\left(\overline{X} - \overline{\overline{X}}\right)$ เราสามารถแสดงได้ดังนี้

การแบ่งกั้นของ $\left(X - \bar{\bar{X}}\right)$

$$\left(X - \bar{\bar{X}}\right) = \left(X - \bar{X}\right) + \left(\bar{X} - \bar{\bar{X}}\right) \qquad \text{..........(9.1)}$$

พิจารณาคะแนนของสามเซต 2,3,4 ($\bar{X}_A = 3$); 5,6,7($\bar{X}_A = 6$);และ 8,9,10; ($\bar{X}_A = 9$); เมื่อ $\bar{\bar{X}} = 6$ สำหรับคะแนน 9 : 9 – 6 = (9-7)+(7-6) = 3

เราสามารถใช้สูตร 9.1 พัฒนา (1) การวัดความแปรปรวนที่สะท้อนเพียงความแปรปรวนที่มีอยู่ประจำตัว (ความแปรปรวนภายในกลุ่ม) และ (2) การวัดที่สะท้อนความแปรปรวนที่มีอยู่ประจำตัวรวมกับผลของการปฏิบัติ (treatment) ถ้านำเสนอ (การเบี่ยงเบนระหว่างกลุ่ม) คำนวณ SS_{total} ผลบวกของความเบี่ยงเบนยกกำลังสองจากค่าเฉลี่ยรวม (grand mean) เราต้องเริ่มจากยกกำลังสองแต่ละค่าของความเบี่ยงเบนจากค่าเฉลี่ยรวมแล้วบวกจำนวนเหล่านี้ เขียนแทนด้วยสัญลักษณ์

$$SS_{total} = \sum^{all\ scores} (X - \bar{\bar{X}})^2$$

จากสูตร 9.1 จะได้

$$\sum^{all\ scores} (X - \bar{\bar{X}})^2 = \sum^{all\ scores} (X - \bar{X})^2 + \sum^{k} n_i (\bar{X}_i - \bar{\bar{X}})^2 \qquad \text{..........(9.2a)}$$

เมื่อ n_i คือจำนวนคะแนนในกลุ่มที่ i , $\bar{X}_i$ คือค่าเฉลี่ยของกลุ่มที่ i และ k เป็นจำนวนของกลุ่ม

จากตัวอย่างทั้งหมด เราจะได้สามชุดที่แตกต่างกันของผลบวกของกำลังสอง

1. $\sum^{all\ scores} (X - \bar{\bar{X}})^2$, ผลบวกทั้งหมดของกำลังสอง เรียก SS_{total}

2. $\sum^{all\ scores}(X - \overline{X})^2$, ผลบวกภายในกลุ่มของกำลังสอง เรียก SS_{within}

3. $\sum^{k} n_i(\overline{X} - \overline{\overline{X}})^2$, ผลบวกระหว่างกลุ่มของกำลังสอง เรียก $SS_{between}$

ดังนั้น สูตร (9.2a) สามารถแสดงดังนี้
การแบ่งส่วนของผลรวมทั้งหมดของกำลังสอง

$$SS_{total} = SS_{within} + SS_{between} \qquad \text{.......... (9.2b)}$$

การแบ่ง SS_{total} เป็นสองส่วนจะอ้างอิงถึงการแบ่งส่วนของผลบวกของกำลังสอง ให้เราตรวจสอบแต่ละส่วนของสูตร 9.2b อย่างใกล้ชิด

1. ผลบวกทั้งหมดของกำลังสอง เป็นการวัดความเบี่ยงเบนทั้งหมดในการนำเสนอข้อมูลโดยปราศจากการพิจารณาถึงกลุ่ม

2. ผลบวกของกำลังสองภายในกลุ่ม

$$SS_{within} = \sum^{all\ scores}(X - \overline{X})^2 = \sum(X_A - \overline{X}_A)^2 + \sum(X_B - \overline{X}_B)^2 + \ldots \qquad \text{..........(9.3)}$$

ต่อจากนี้เราจะเขียนแทน SS_{within} ด้วย SS_w สิ่งนี้เป็นการวัดของการเบี่ยงเบนภายในกลุ่มของคะแนนแต่ละตัว ความเสรีของความแตกต่างใดๆระหว่างค่าเฉลี่ยของกลุ่ม มันเป็นผลสะท้อนโดยตรงของการเบี่ยงเบนที่มีอยู่ดั้งเดิม(inherent) อิสระจากอิทธิพลของผลการปฏิบัติที่แตกต่างกัน การประมาณควรทำโดยดำเนินการผลบวกของกำลังสองสำหรับหนึ่งกลุ่มใดๆ อย่างไรก็ตามในการตั้งสมมุติฐานว่าความแปรปรวนของประชากรเหมือนกันสำหรับทุกกลุ่ม ($\sigma_A^2 = \sigma_B^2 = \sigma_C^2 = \ldots$ข้อสมมุติฐานของความเป็นเอกพันธ์ของความแปรปรวน)เราอาจได้การประมาณที่ดีกว่าโดยเชื่อมข้อมูลจากหลายๆกลุ่ม ความแปรปรวนที่ถูกตั้งสมมุติฐานที่เป็นค่าร่วมกันกับทุกประชากรคือ σ^2 (ซึ่งไม่มีตัวห้อย)

3. ผลบวกของกำลังสองระหว่างกลุ่ม

$$SS_{between} = \sum^{k} n_i(\overline{X}_i - \overline{\overline{X}})^2 = n_A(\overline{X}_A - \overline{\overline{X}})^2 + n_B(\overline{X}_B - \overline{\overline{X}})^2 + \ldots \qquad \text{.......... (9.4)}$$

ต่อจากนี้เราจะเขียนแทน $SS_{between}$ ด้วย SS_{bet} เป็นการวัดความเบี่ยงเบนของค่าเฉลี่ยกลุ่มตัวอย่างระหว่างกลุ่ม มันสะท้อนการเบี่ยงเบนที่มีอยู่ประจำตัว(อิสระจากอิทธิพล

ของผลการปฏิบัติ) บวกผลความแตกต่างของการปฏิบัติใดๆ (ความเบี่ยงเบนซึ่งเป็นผลจากความแตกต่างระหว่างค่าเฉลี่ยประชากร)

4. องศาเสรี (degrees of freedom)

การทดสอบ H_0 ในการวิเคราะห์ความแปรปรวนทางเดียวเกี่ยวข้องกับการเปรียบเทียบสองการประมาณที่แบ่งแยกกันสำหรับประชากร อันหนึ่งขึ้นกับการเบี่ยงเบนภายในกลุ่มระหว่างคะแนนแต่ละตัวเทียบกับคะแนนเฉลี่ยของกลุ่ม และอีกอันหนึ่งเป็นความเบี่ยงเบนระหว่างค่าเฉลี่ยของกลุ่มตัวอย่าง ซึ่งเราได้ประมาณความแปรปรวนของประชากรที่ไม่ลำเอียง (unbiased) ไปแล้ว เราหารผลบวกของกำลังสองด้วยองศาเสรีที่เกี่ยวข้องกับผลบวกของกำลังสอง ความสัมพันธ์ทั่วไปนี้สรุปได้ด้วยสูตร

$$s^2 = \frac{SS}{df}$$

ระลึกว่าองศาเสรีสำหรับการประมาณความแปรปรวนขึ้นกับตัวอย่างเชิงเดี่ยวคือ n-1 และองศาเสรีสำหรับการประมาณรวมของความแปรปรวนใช้ในการทดสอบ t สำหรับสองค่าเฉลี่ยที่เป็นอิสระต่อกันคือ $(n_1-1)+(n_2-1)$

องศาเสรีที่เกี่ยวข้องกับ SS_{total} คือคุณอาจคาดหวังว่าเป็น $n_{total}-1$ เพราะว่ามี n-1 องศาเสรีที่เกี่ยวข้องกับความเบี่ยงเบนเทียบกับค่าเฉลี่ยเชิงเดี่ยว องศาเสรีสำหรับ SS_w คือ $(n_A-1)+(n_B-1)+\ldots = n_A+n_B+\ldots+n_k-k$ หรือ

$$df_w = n_{total} - k$$

เมื่อ n_{total} เป็นผลรวมจำนวนของกรณีและ k เป็นจำนวนกลุ่ม สำหรับ SS_{bet} จะมีเพียงการเบี่ยงเบนมากจากค่าเฉลี่ยรวมเช่นเดียวกับมีค่าเฉลี่ยกลุ่มตัวอย่าง(หรือจำนวนกลุ่ม) ดังนั้น

$$df_{bet} = k-1$$

สำหรับการวิเคราะห์ความแปรปรวนใดๆ จำนวนขององศาเสรีเชื่อมสัมพันธ์กับ SS_w บวกจำนวนองศาเสรีเชื่อมสัมพันธ์กับ SS_{bet} เท่ากับจำนวนขององศาเสรีเชื่อมสัมพันธ์กับ SS_{total}

$$df_{total} = df_w + df_{bet}$$

5 การประมาณความแปรปรวนและอัตราส่วน F (Variance estimates and the F ratio)

ถ้าเราหาร SS_w และ SS_{bet} ด้วยองศาเสรีของมันตามลำดับ เราจะได้สองการประมาณความแปรปรวนที่ต้องทดสอบ $H_0: \quad \mu_A = \mu_B = \mu_C$ การประมาณความแปรปรวนภายในกลุ่ม (s_w^2) และการประมาณความแปรปรวนระหว่างกลุ่ม (s_{bet}^2) การประมาณความแปรปรวนเหล่านี้คือ

การประมาณความแปรปรวนภายในกลุ่ม

$$s_w^2 = \frac{SS_w}{df_w} \xrightarrow{\text{การประมาณ}} \sigma^2 \text{ (ความแปรปรวนประจำตัว)} \qquad \text{..........(9.5)}$$

การประมาณความแปรปรวนระหว่างกลุ่ม

$$s_{bet}^2 = \frac{SS_{bet}}{df_{bet}} \xrightarrow{\text{การประมาณ}} \sigma^2 + \text{ผลการปฏิบัติ} \qquad \text{..........(9.6)}$$

บ่อยครั้ง นักคณิตศาสตร์จะอ้างอิงถึง s_w^2 กำลังสองค่าเฉลี่ยภายในกลุ่ม(หรือบางครั้งความคลาดเคลื่อนกำลังสองค่าเฉลี่ย(the mean square error) และ s_{bet}^2 กำลังสองค่าเฉลี่ยระหว่างกลุ่ม โดยใช้สัญลักษณ์ MS_w และ MS_{bet} ตามลำดับ

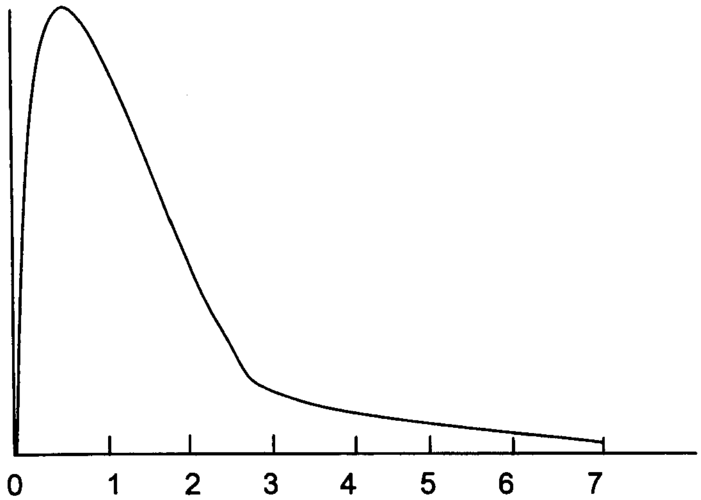
ถ้า H_0 เป็นจริง(นั่นคือไม่มีความแตกต่างจากผลการปฏิบัติ) ค่าเฉลี่ยของกลุ่มตัวอย่างจะใกล้เคียง $\bar{\bar{X}}$ และ s_{bet}^2 จะเป็นการประมาณที่ไม่ลำเอียงของความแปรปรวนที่มีอยู่

ประจำตัว σ^2 เพราะฉะนั้นการประมาณปริมาณเดียวกันเช่นการประมาณโดย s_w^2 และทั้งสองควรจะเท่ากันภายในขีดจำกัดของการเบี่ยงเบนการสุ่ม อีกด้านหนึ่ง ถ้า H_0 เป็นเท็จ (ถ้ามีผลเกิดจากการปฏิบัติ) ผลบวกของกำลังสองของการเบี่ยงเบนของ $\overline{X}$ เทียบกับ $\overline{\overline{X}}$ จะมีค่าใหญ่กว่าและ s_{bet}^2 จะใหญ่กว่า s_w^2 ซึ่งมันเกินกว่าลิมิตของการเบี่ยงเบนการสุ่ม การวิเคราะห์ที่สมบูรณ์ของความแปรปรวน เราต้องการวิธีของการเปรียบเทียบ s_{bet}^2 กับ s_w^2 ซึ่งเรียกว่าอัตราส่วน F ชื่อทางสถิติเพื่อเป็นเกียรติกับ Ronald Fisher อัตราส่วน F เป็นอัตราส่วนของสองการประมาณความแปรปรวนที่เป็นอิสระต่อกัน

อัตราส่วน F สำหรับการวิเคราะห์ความแปรปรวนทางเดียว

$$F = \frac{s_{bet}^2}{s_w^2} \qquad \text{..........(9.7)}$$

ถ้า H_0 เป็นจริง s_{bet}^2 และ s_w^2 ทั้งสองจะเป็นการประมาณของความแปรปรวนที่มีอยู่ประจำตัวเท่านั้น และอัตราส่วนจะมีค่าประมาณ 1 ถ้า H_0 เป็นเท็จ s_{bet}^2 จะเป็นการประมาณของความแปรปรวนที่มีอยู่ประจำตัวบวกกับความแตกต่างจากผลการปฏิบัติและอัตราส่วนจะไกลเกิน 1 ความแตกต่างผลการปฏิบัติที่มากกว่าจะทำให้ค่า F มีค่าไกลเกิน 1 เมื่อ H_0 เป็นจริงและยอมรับสมมุติฐาน อัตราส่วน F เป็นไปตามทฤษฎี การแจกแจง F ที่นำเสนอในตาราง C ในภาคผนวก คล้ายการแจกแจง t การแจกแจง F แท้จริงมีโค้งที่ขึ้นกับองศาเสรี ในการพิจารณาค่าในตารางของ F สำหรับเปรียบเทียบกับค่าในการศึกษา เราต้องพิจารณาสองค่าสำหรับ df ที่เกี่ยวข้องกับตัวเศษ s_{bet}^2 และที่เกี่ยวข้องกับตัวส่วน s_w^2 รูป 9.3 แสดงการแจกแจงของ F สำหรับองศาเสรี 4 ในตัวเศษ และองศาเสรี 20 สำหรับตัวส่วน



F Ratio รูป 9.3

การแจกแจงนี้เบ้ทางบวก (positively skewed) สังเกตว่า ถ้าการประมาณของ σ^2 ในตัวเศษน้อยกว่าในตัวส่วน F จะน้อยกว่า 1 แต่ไม่น้อยกว่าศูนย์ (การประมาณความแปรปรวนไม่เป็นลบ) ถ้าการประมาณในตัวเศษมากกว่าในตัวส่วน อัตราส่วน F จะมากกว่า 1 สมมุติฐานว่างของสมการของค่าเฉลี่ยประชากรจะถูกปฏิเสธเพียงถ้าการคำนวณค่า F ใหญ่กว่าการคาดหมายโดยตลอดการชักตัวอย่างการสุ่มถ้าสมมุติฐานเป็นจริง สิ่งที่ตามมาคือบริเวณของการปฏิเสธคือทั้งหมดในหางด้านบนของการแจกแจง F (ระลึกว่า การทดสอบ F ไม่มีทิศทาง) ค่า F น้อยกว่า 1 บ่งชี้ว่าไม่พบหนึ่งในสมมุติฐานต่างๆ และ/หรือมีความคลาดเคลื่อนจากการสุ่ม

ตารางสรุป เพื่อความสะดวกในการนำเสนอแหล่งของความเบี่ยงเบน ตาราง 9.1 แสดงรูปทั่วไปของการวิเคราะห์ความแปรปรวนทางเดียว สดมภ์แรกแสดงแหล่งของความแปรปรวน สังเกตว่าความเบี่ยงเบนระหว่างกลุ่มแจงความเบี่ยงเบนภายในกลุ่มก่อน ดังนั้นในการกำหนดอัตราส่วน F ผลบวกของกำลังสองจะแจงในสดมภ์ที่สอง และองศาเสรีในสดมภ์ที่สาม เมื่อตรวจสอบการทำงาน $SS_{bet}+SS_w$ ควรเท่ากับ SS_{total} และ $df_{bet}+df_w$ ควรเท่ากับ df_{total} ตัวเศษและตัวส่วนของอัตราส่วน F ได้จากการหาร SS ด้วย df และแสดงในสดมภ์ที่สี่ สังเกตว่า $s^2_{bet}+s^2_w$ ไม่เท่ากับ s^2_{total} ซึ่งตาราง 9.1 ไม่ได้แสดง s^2_{total}

ตาราง 9.1 ตารางสรุปของการวิเคราะห์ความแปรปรวนทางเดียว (One-Way ANOVA) แสดงการแบ่งของผลบวกของกำลังสอง องศาเสรี การประมาณความแปรปรวน และ F

SOURCE	SS	df	s^2	F
Between groups	Formula 9.4	k-1	$\frac{SS_{bet}}{df_{bet}}$	$\frac{s^2_{bet}}{s^2_w}$
Within groups	Formula 9.3	n_{total} - k	$\frac{SS_w}{df_w}$	
Total	Fomula 9.2a	n_{total} - 1		

6. ตัวอย่างการวิเคราะห์ความแปรปรวน

ในการแสดงการคำนวณสำหรับการวิเคราะห์ความแปรปรวนทางเดียว เราจะใช้สมมุติตัวอย่างที่มีคะแนนเพียงไม่กี่ตัว สมมุติว่าผู้อำนวยการสถาบันแห่งหนึ่งต้องการเปรียบเทียบว่านักศึกษาปีหนึ่งในสถาบันตอบสนองต่อการสอนวิธีใหม่สามวิธีอย่างไร ในตอนเริ่มต้น เขาสุ่มนักศึกษาปีหนึ่งมา 15 คน และสุ่มลงสามกลุ่ม แต่ละกลุ่มจะได้รับวิธีการสอนตามวิธีที่กำหนดในแต่ละกลุ่มซึ่งแตกต่างกัน เป็นเวลาเก้าสัปดาห์ และทดสอบ โดยกำหนดสมมุติฐานเป็น

$$H_0:\ \mu_A = \mu_B = \mu_C$$

ถ้ากลุ่มใดกลุ่มหนึ่งแตกต่างจากกลุ่มอื่นหรือทั้งสามกลุ่มแตกต่างกัน เขาจะปฏิเสธ H_0 และยอมรับ H_A

ในขณะทำการทดลองนักศึกษาคนหนึ่งในกลุ่ม C หยุดเรียนก่อนการทดลองเสร็จสิ้น ดังนั้นจะมีสองกลุ่มที่มีนักศึกษากลุ่มละ 5 คน และมีกลุ่มหนึ่งมีนักศึกษา 4 คน เมื่อเสร็จสิ้นการทดลองได้มีการสอบวัดผลสัมฤทธิ์ ปรากฏผลดังนี้

กลุ่ม A	กลุ่ม B	กลุ่ม C
8	13	4
11	11	6
8	13	5
9	10	7
8	13	

หาค่าเฉลี่ยของกลุ่มตัวอย่างแต่ละกลุ่มและค่าเฉลี่ยรวมได้ดังนี้

$$\overline{X}_A = \frac{44}{5} = 8.8, \overline{X}_B = \frac{60}{5} = 12, \ \overline{X}_C = \frac{22}{4} = 5.5, \overline{\overline{X}} = \frac{126}{14} = 9$$

ขั้นต่อไปหา SS_w และ SS_{bet} :

$$SS_w = \sum^{\substack{all \\ scores}} (X - \overline{X})^2$$

$$= \sum (X_A - \overline{X}_A)^2 + \sum (X_B - \overline{X}_B)^2 + \sum (X_C - \overline{X}_C)^2$$

$$= \left[(8-8.8)^2 + (11-8.8)^2 + (8-8.8)^2 + (9-8.8)^2 + (8-8.8)^2\right]$$

$$+ \left[(13-12)^2 + (11-12)^2 + (13-12)^2 + (10-12)^2 + (13-12)^2\right]$$

$$+ \left[(4-5.5)^2 + (6-5.5)^2 + (5-5.5)^2 + (7-5.5)^2\right]$$

$$= \left[0.64 + 4.84 + 0.64 + 0.04 + 0.64\right] + \left[1+1+1+4+1\right]$$

$$+ \left[2.25 + 0.25 + 0.25 + 2.25\right]$$

$$= 6.8 + 8 + 5$$

$$= 19.8$$

SS_{bet} = $\sum^{k} n_i(\overline{X}_i - \overline{\overline{X}})^2$

= 5(8.8-9)^2 + 5(12-9)^2 + 4(5.5-9)^2

= 0.2 + 45 + 49

= 94.2

แม้ว่า SS_{total} จะไม่ได้นำมาใช้ในการหาอัตราส่วน F แต่สามารถใช้ตรวจสอบความถูกต้องได้ ดังนั้นในที่นี้จะคำนวณหา SS_{total} ด้วย

SS_{total} = $\sum(X - \overline{\overline{X}})^2$

= (8-9)^2 + (11-9)^2 + (8-9)^2 + (9-9)^2 + (8-9)^2 + (13-9)^2 + (11-9)^2 + (13-9)^2 +(10-9)^2 +(13-9)^2 + (4 -9)^2 + (6-9)^2 + (5-9)^2 + (7-9)^2

= 1+4+1+0+1+16+4+16+1+16+25+9+16+4

=114

ตาราง 9.2 ตารางสรุปการวิเคราะห์ความแปรปรวนสำหรับทดสอบสมมุติฐาน $H_0 : \mu_A = \mu_B = \mu_C$ สำหรับคะแนนผลสัมฤทธิ์

SOURCE	SS	df	s^2	F
Between groups	94.2	2	47.1	26.17
Within group	19.8	11	1.8	
Total				

เพราะว่า $SS_{total} = SS_w + SS_{bet}$ เราจึงแน่ใจได้ว่าการคิดคำนวณถูกต้อง

คิดคำนวณค่าต่าง ๆเพื่อเติมค่าในตาราง 17.2 ดังนี้

คำนวณองศาเสรี

df_{bet} = k-1 = 3-1 = 2

df_w = n_{total} - k = 14 – 3 = 11

df_{total} = n_{total} - 1 = 14 –1 = 13

จะเห็นว่า $df_{total} = df_{bet} + df_{w}$

คำนวณหา s^2_{bet} และ s^2_{w}

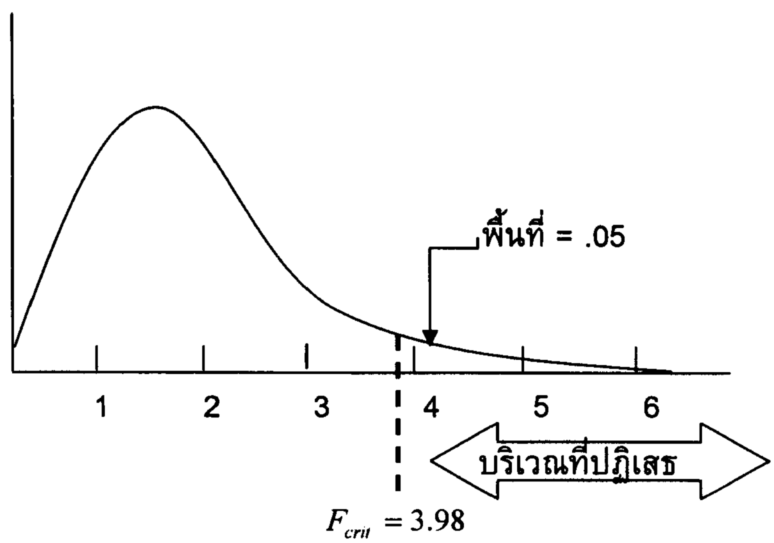
$$s^2_{bet} = \frac{SS_{bet}}{df_{bet}} = \frac{94.2}{2} = 47.1$$

$$s^2_{w} = \frac{SS_{w}}{df_{w}} = \frac{19.8}{11} = 1.8$$

หาอัตราส่วน F

$$F = \frac{s^2_{bet}}{s^2_{w}} = \frac{47.1}{1.8} = 26.17$$

ในตัวอย่างนี้ $df_{bet} = 2$ และ $df_{w} = 11$ เปิดตาราง C ในภาคผนวก ที่เป็นส่วนตัดของ 2 df สำหรับตัวเศษ(numerator) และ 11 df สำหรับตัวส่วน (denumerator) ถ้าผลการสอนใช้ระดับนัยสำคัญที่ระดับ .05 ค่าวิกฤติของ F คือ 3.98(หรือ7.20 สำหรับระดับนัยสำคัญ .01 ถ้า H_0 เป็นจริง เราจะได้อัตราส่วน F มากกว่า 3.98 เพียง 5% ของการชักการสุ่ม เนื่องจากเราได้ F = 26.17 ตกนอกเหนือค่าวิกฤติ ดังนั้นเราจะปฏิเสธ H_0 การแจกแจง F สำหรับ 2 และ 11 องศาเสรี และบริเวณปฏิเสธแสดงในภาพ 9.4



รูป 9.4 การแจกแจงการสุ่มของ F สำหรับ 2 และ 11 องศาเสรีของ $\alpha = .05$

7. การวิเคราะห์ความแปรปรวนจากสูตรคะแนนดิบ

กลุ่ม A $n_A = 5$

คำนวณ $\sum X_A =$ 8+11+8+9+8 =44

$\sum X_A^2 =$ 64+121+64+81+64 = 394

กลุ่ม B $n_B = 5$

คำนวณ $\sum X_B =$ 13+11+13+10+13 = 60

$\sum X_B^2 =$ 169+121+169+100+169 = 728

กลุ่ม C $n_C = 4$

คำนวณ $\sum X_C =$ 4+6+5+7 = 22

$\sum X_C^2 =$ 16+36+25+49 = 126

ขั้นต่อไปบวกค่าของ $\sum X, \sum X^2$ และรวมค่าของแต่ละกลุ่มจะได้

$$\sum^{\substack{all\\scores}} X = \sum X_A + \sum X_B + \sum X_C = 44+60+22 = 126$$

$$\sum^{\substack{aii\\scores}} X^2 = \sum X_A^2 + \sum X_B^2 + \sum X_C^2 = 394+728+126 = 1248$$

$$n_{total} = n_A + n_B + n_C$$

$$= 5+5+4$$

$$= 14$$

หา SS_w, SS_{bet} และ SS_{total} ดังนี้

สูตรคะแนนดิบ ผลบวกของกำลังสองภายในกลุ่ม

$$SS_w = \sum^{\substack{aii\\scores}} X^2 - \left[\frac{\left(\sum X_A\right)^2}{n_A} + \frac{\left(\sum X_B\right)^2}{n_B} + \ldots\right]$$

ดังนั้น

$$SS_w = 1248 - \left[\frac{(44)^2}{5} + \frac{(60)^2}{5} + \frac{(22)^2}{4}\right]$$

$$= 1248 - (387.2 + 720 + 121)$$

$$= 19.8$$

สูตรคะแนนดิบ ผลบวกของกำลังสองระหว่างกลุ่ม

$$SS_{bet} = \left[\frac{\left(\sum X_A\right)^2}{n_A} + \frac{\left(\sum X_B\right)^2}{n_B} + \ldots\right] - \frac{\left(\sum^{all\ score} X\right)}{n_{total}} \qquad \text{.......... (9.9)}$$

ดังนั้น

$$SS_{bet} = \left[\frac{(44)^2}{5} + \frac{(60)^2}{5} + \frac{(22)^2}{4}\right] - \frac{(126)^2}{14}$$

$$= 1228.2 - 1134$$

$$= 94.2$$

สูตรคะแนนดิบ ผลบวกของกำลังสองของทั้งหมด

$$SS_{total} = \sum^{all\ scores} X^2 - \frac{\left(\sum^{all\ scores} X\right)^2}{n_{total}} \qquad \text{.......... (9.10)}$$

ดังนั้น

$$SS_{total} = 1248 - \frac{(126)^2}{14}$$

$$= 1248 - 1134$$

$$= 114$$

สังเกตว่าค่าสำหรับ SS_w , SS_{bet} และ SS_{total} เหมือนกับที่คำนวณมาแล้ว

8. การเปรียบเทียบ t และ F

การวิเคราะห์ความแปรปรวนสามารถใช้ทดสอบสมมุติฐาน H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ ซึ่งมีสองกลุ่มหรือมากกว่า ดังนั้นเราสามารถใช้ ANOVA แทน t ในกลุ่มตัวอย่างสองกลุ่มที่เป็นอิสระต่อกัน ตามข้อเท็จจริงสำหรับกลุ่มตัวอย่างที่มีสองกลุ่ม F = t^2 ตัวอย่างเช่น ถ้านักศึกษาคำนวณโดยใช้ t ได้ 2.074 จะหาได้ว่า F = $(2.074)^2$ = 4.30

9. ข้อตกลงที่เกี่ยวกับการวิเคราะห์ความแปรปรวน

เพราะว่า F = t^2 เราจึงควรรู้ว่าการทดสอบ F สำหรับความแปรปรวนทางเดียวเหมือนกับการทดสอบ t สำหรับกลุ่มตัวอย่างที่เป็นอิสระต่อกัน ซึ่งเราจะกล่าวอีกครั้งดังนี้

1. ประชากรเป็นการแจกแจงปกติ
2. ความแปรปรวนของประชากรต่าง ๆเหมือนกัน(ความเป็นเอกพันธ์ของความแปรปรวน)
3. เลือกสมาชิกที่ประกอบขึ้นเป็นกลุ่มตัวอย่างที่เป็นอิสระต่อกัน
4. ตัวอย่างที่ถูกชักจากการสุ่มด้วยการแทนที่

10 การเปรียบเทียบ POST HOC

สมมุติว่าผลการทดลองวิธีสอนของสามกลุ่มที่กล่าวแล้ว ได้ อัตราส่วน F เล็กกว่าค่าวิกฤตของ 3.98 เราจะสรุปว่าข้อมูลของเราไม่สนับสนุนผลการสอนที่แตกต่างกันนั่นคือ

ไม่มีความแตกต่างในการสอนวิธีต่าง ๆ แต่กรณีของเราไม่เป็นเช่นนี้ เนื่องจากเราได้อัตราส่วน F เป็น 26.17 ทำให้ราต้องปฏิเสธ H_0: $\mu_A = \mu_B = \mu_C$ และยอมรับสมมุติฐานแย้งว่าค่าเฉลี่ยประชากรแตกต่างกันในบางวิธี แต่วิธีไหนที่แตกต่างกัน อาจทั้งสามวิธีแตกต่างกันหรือมีสองวิธีในวิธีเหล่านั้นที่แตกต่างกัน เพื่อจะตอบคำถามนี้ เราจะเปรียบเทียบผลทางสถิติที่เกี่ยวข้องกับค่าเฉลี่ยของกลุ่ม

ในการทดสอบเราใช้การเปรียบเทียบที่เรียกว่า post hoc comparisons หรือ posteriori (ภาษาละตินซึ่งหมายถึง "what come later") comparisons ซึ่งแทนที่เราจะใช้การแจกแจงการสุ่มที่เปรียบเทียบค่าเฉลี่ยเฉพาะสองกลุ่มตัวอย่าง(เช่นที่ทำกับt) การทดสอง post hoc จะจัดหาการแจกแจงการสุ่มที่เปรียบเทียบค่าเฉลี่ยของหลายกลุ่มตัวอย่าง กล่าวสั้นๆ การทดสอบ post hoc ป้องกันเราจากการทำให้เกิดความคลาดเคลื่อนแบบที่ I โดยทำให้ความแตกต่างมากเกินไป (ระหว่างค่าเฉลี่ยของกลุ่มตัวอย่าง) ก่อนที่เราจะประกาศว่าความแตกต่างมีนัยสำคัญทางสถิติ มีการใช้การทดสอบpost hoc จากหลายแบบ เช่น การทดสอบที่เรียกว่า Duncan's multiple-range test , the Newman-Keuls test, Tukey's HSD test, หรือ the Scheffe test เราจะใช้การทดสอบแบบใดแบบหนึ่งที่กล่าวมานี้ แต่ในที่นี้จะกล่าวเพียงแบบเดียวคือTukey's HSD test การทดสอบจะเกี่ยวข้องกับการตรวจสอบค่าวิกฤต HSD สำหรับข้อมูล แม้ว่าการทดสอบสามารถเปรียบเทียบกันกรณีเฉพาะเจาะจง แต่ส่วนใหญ่จะสืบสวนสอบสวนโดยเปรียบเทียบทีละคู่ สมมุติฐานของการเท่ากันของค่าเฉลี่ยของประชากรจะถูกปฏิเสธสำหรับคู่ใด ๆของกลุ่มตัวอย่างซึ่งความแตกต่างระหว่าง $\overline{X}$ มีค่ามากหรือมากกว่าค่าวิกฤต HSD

เราจะตรวจสอบค่าวิกฤต HSD โดยใช้สูตรต่อไปนี้

ค่าวิกฤต HSD สำหรับการทดสอบ HSD ของ Tukey

$$\text{HSD} = q\sqrt{\frac{s_w^2}{n}} \qquad \text{..........} \ (9.11)$$

เมื่อ s_w^2 คือค่าประมาณความแปรปรวนภายในกลุ่ม

n คือจำนวนกรณีในแต่ละกลุ่ม

q คือค่าสถิติใหม่ที่เรียก studentized range statistic ขึ้นกับการเปรียบเทียบการแจกแจงการสุ่มของค่าเฉลี่ยที่มากกว่าสองกลุ่ม)

ค่าของ q ขึ้นอยู่กับระดับนัยสำคัญ (.05 หรือ.01) จำนวนของกลุ่มตัวอย่างที่จะเปรียบเทียบ และ df_w ในการหา q ให้ดูตาราง F ในภาคผนวก ถ้าจำนวนของกรณีไม่เหมือนกันสำหรับทุกกลุ่ม คุณต้องใช้ค่าเฉลี่ยฮาร์โมนิก (harmonic mean) ,$\tilde{n}$,ในตำแหน่งของ n ซึ่งค่าเฉลี่ยฮาร์โมนิกมีวิธีคำนวณดังนี้

$$\tilde{n} = \frac{k}{(1/n_A) + (1/n_B) + \ldots + (1/n_k)} \qquad \text{..........(9.12)}$$

เมื่อ: k = จำนวนของกลุ่ม

n_k = จำนวนของกรณีในกลุ่มที่ k

เพื่อแสดงการใช้การทดสอบ HSD ของ Turkey เราจะกลับไปใช้ปัญหาในหัวข้อ 6 ซึ่งเราได้ว่ามันแตกต่างอย่างมีนัยสำคัญระหว่างวิธีการสอนสามวิธี เราต้องการตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่ม

ในตัวอย่างนั้น df_w = 11, k = 3, และ α = .05 เมื่อเปิดตาราง F ด้วยค่าเหล่านี้จะได้ q = 3.82 เพราะว่าขนาดของกลุ่มตัวอย่างไม่เท่ากัน เราต้องคำนวณค่าเฉลี่ยฮาร์โมนิก

$$\tilde{n} = \frac{k}{(1/n_A) + (1/n_B) + (1/n_k)}$$

$$= \frac{3}{(1/5) + (1/5) + (1/4)} = \frac{3}{.65} = 4.62$$

เรามี s_w^2 = 1.8 ค่าวิกฤต HSD คือ

$$\text{HSD} = 3.82\sqrt{\frac{1.8}{4.62}}$$

$$= 0.516$$

ในตอนนี้เราสามารถคำนวณความแตกต่างสำหรับทุกคู่ที่เป็นไปได้ของค่าเฉลี่ยกลุ่ม

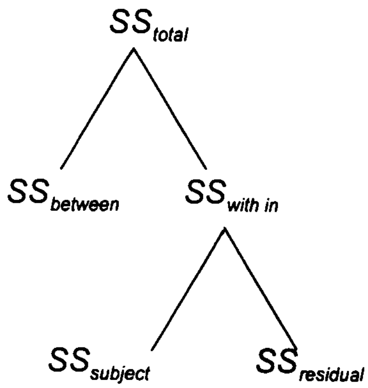
ตาราง7.3 แสดงแต่ละสมาชิกเป็นความแตกต่างระหว่างค่าเฉลี่ยที่แจงบนด้านซ้ายและที่ด้านบน

	$\bar{X}_A=8.8$	$\bar{X}_B=12$	$\bar{X}_C=5.5$
$\bar{X}_A=8.8$	0	−3.2	3.3
$\bar{X}_B=12$		0	6.5
$\bar{X}_C=5.5$			0

ในตัวอย่างนี้ ความแตกต่างระหว่างทุกคู่ของค่าเฉลี่ยของกลุ่มตัวอย่างมากกว่า 2.28 ดังนั้นจึงมีนัยสำคัญที่ระดับ.05 เราสามารถสรุปว่าวิธีสอนแบบ B ดีกว่าวิธีสอนแบบ A และทั้งสองวิธีดีกว่าวิธีสอนแบบ C

11. การวิเคราะห์ความแปรปรวนสำหรับการวัดซ้ำ (Analysis of Variance for Repeated Measure)

ตรรกะของการวิเคราะห์ความแปรปรวนทางเดียวสำหรับการวัดซ้ำเหมือนกับการวัดที่เป็นอิสระต่อกัน เราแบ่งกั้นการเบี่ยงเบนทั้งหมดของข้อมูลเป็นสองส่วนคือภายในกลุ่มและระหว่างกลุ่ม ในการวิเคราะห์ความแปรปรวนสำหรับการวัดซ้ำ SS_w จะถูกแบ่งกั้นเป็นสองส่วน (1) $SS_{subjects}$ เป็นการวัดการเปลี่ยนแปลงเกี่ยวกับความแตกต่างระหว่างบุคคล และ (2) $SS_{residual}$ วัดความเบี่ยงเบนเกี่ยวกับความคลาดเคลื่อนจากการสุ่ม รูป 9.5 แสดงการแบ่งกั้นสำหรับการวิเคราะห์ความแปรปรวนสำหรับการวัดซ้ำ



รูป 9.5 การแบ่งกั้นของผลบวกของกำลังสองสำหรับการวิเคราะห์ความแปรปรวนสำหรับการวัดซ้ำ

เราสามารถวัดความเบี่ยงเบนที่เกี่ยวกับความแตกต่างระหว่างบุคคล แต่ละสมาชิกจะถูกทดสอบภายใต้เงื่อนไขทั้งหมด ซึ่งจะช่วยให้เราสามารถรายงานสมาชิกทั้งหมด (โดยรวมคะแนนทั้งหมดของเขาเข้าด้วยกัน) ซึ่งเราจะเปรียบเทียบกับสมาชิกอื่นๆของคะแนนทั้งหมด ถ้าคะแนนสองคะแนนของคะแนนทั้งหมดแตกต่างกัน เราจะถือเอาคะแนนความแตกต่าง ในกรณีที่คะแนนสองคะแนนใดๆทั้งหมดเหมือนกันยกเว้นแตกต่างภายใต้เงื่อนไข เราจะถือเอาความคลาดเคลื่อนจากการชัก (random)

สูตรสำหรับ SS_{total} เหมือนกับการวิเคราะห์ความแปรปรวนทางเดียวกับกลุ่มที่เป็นอิสระต่อกัน (สูตร 9.9)

$$SS_{total} = \sum^{\substack{all \\ score}} (X - \bar{\bar{X}})^2 \qquad \text{..........(9.13)}$$

สูตรสำหรับ SS_{bet} ก็เหมือนกันด้วย เพราะว่าแต่ละกลุ่มมีจำนวน subject เหมือนกันในแบบแผนการวัดซ้ำ เราสามารถดัดแปลงสูตรคะแนนดิบสำหรับ SS_{bet} (สูตร 9.9)

SS_{bet} : วิธี คะแนนความเบี่ยงเบน (Deviation-Scores Method)

$$SS_{bet} = \sum n_i(\overline{X}_i - \overline{\overline{X}})^2 \qquad \text{.......... (9.14a)}$$

SS_{bet} : วิธีคะแนนดิบสำหรับการวัดซ้ำ

$$SS_{bet} = \frac{(\sum X_A)^2 + (\sum X_B)^2 + \ldots}{n_k} - \frac{(\sum^{all\ score} X)}{n_{total}} \qquad \text{.......... (9.14b)}$$

คำนวณ $SS_{subjects}$ โดยใช้สูตรใดสูตรหนึ่งต่อไปนี้

$$SS_{subj} = \sum k(\overline{X}_{subj} - \overline{\overline{X}})^2 \qquad \text{.......... (9.15a)}$$

เมื่อ : $\overline{X}_{subj}$ คือค่าเฉลี่ยของคะแนน subject กับทุกเงื่อนไข

k คือจำนวนเงื่อนไข

$SS_{subjects}$: วิธีคะแนนดิบ

เมื่อ k คือจำนวนเงื่อนไข

$$SS_{subj} = \frac{\sum X^2_{subj}}{k} - \frac{(\sum^{all\ score} X)^2}{n_{total}} \qquad \text{.......... (9.15b)}$$

เราหา $SS_{residual}$ โดยการลบ

$$SS_{resid} = SS_{total} - SS_{bet} - SS_{subj} \qquad \text{.......... (9.16)}$$

องศาเสรีสำหรับ SS_{subj} คือจำนวนของสมาชิก ลบด้วยหนึ่ง สำหรับ SS_{resid} องศาเสรีคือ $(df_{subj})(df_{bet})$ ในการตรวจสอบว่าถ้ามีผลจากการปฏิบัติ เราคำนวณอัตราส่วน F โดยหาร s^2_{bet} ด้วย s^2_{resid}

$$F = \frac{s^2_{bet}}{s^2_{resid}} = \frac{random\,error + treatment\,effect}{random\,error} \quad \text{.......... (9.17)}$$

ตาราง 9.4 สรุปตารางของการวิเคราะห์ความแปรปรวนทางเดียวสำหรับการวัดซ้ำ

SOURCE	SS	df	s^2	F
Subject	สูตร 9.15aหรือ9.15b	$n_{total} - 1$	$\frac{SS_{subj}}{df_{subj}}$	-
Between groups	สูตร 9.14aหรือ9.14b	k-1	$\frac{SS_{bet}}{df_{bet}}$	$\frac{s^2_{bet}}{s^2_{resid}}$
Residual	สูตร 9.16	$(df_{subj})(df_{bet})$		
Total	สูตร 9.2a	$n_{total} - 1$		

ตาราง 9.4 ให้รูปแบบทั่วไปของตารางสรุปสำหรับ ANOVA สำหรับการวัดซ้ำ ANOVA ทางเดียวสำหรับการวัดซ้ำเป็นแบบแผนการวิจัยที่มีประสิทธิภาพมากกว่า ANOVA ทางเดียวสำหรับการวัดที่เป็นอิสระต่อกัน เพราะว่ามันเปลี่ยนการเบี่ยงเบนไปขึ้นกับความแตกต่างของแต่ละรายตัว (individual) (s^2_{bet} หารด้วย s^2_{resid} เล็กกว่า s^2_w)

ในการสาธิต ANOVA สำหรับการวัดซ้ำ เมื่อเรามองย้อนกลับไปถึงตัวอย่างสมมุติฐาน ในกรณีของตัวแปรอิสระต่อกัน เราต้องการทดสอบสามระดับของความอิสระ: ศูนย์ (Z) พอประมาณ (M) สูง (H) ของปริมาณยา ในการทดลองยาขนานใหม่สำหรับเด็กกระตือรือร้นเกินปกติ (hyperactive) ในกรณีของตัวแปรไม่เป็นอิสระต่อกัน เราจะเลือกคะแนนของการอ่านแบบทดสอบเพื่อดูความตั้งใจในเวลา 30 นาที

สมมุติฐานว่างและสมมุติฐานแย้งคือ

H_0: $\mu_Z = \mu_M = \mu_H$

H_A: not H_0

เราจะสุ่มเลือกนักเรียนชั้นประถมศึกษาปีที่ 3 จำนวน 12 คน จากโรงเรียนในเขตการศึกษาหนึ่งและทำการทดสอบ มีนักเรียนจำนวน สี่คนแรกได้ขนาดของยาเป็นศูนย์ในตอนแรก แล้วตามด้วยพอประมาณ แล้วเป็นสูง ในเวลาหนึ่งสัปดาห์ระหว่างทดสอบ สี่คนต่อไปได้รับยาตามลำดับเป็น M, H, Z และสี่คนที่เหลือรับยาตามลำดับเป็น H, Z, M ตาราง 9.5 ให้ผลสำหรับการทดลอง (ข้อมูลสมมุติ) คำนวณผลบวกของกำลังสองได้ดังการคำนวณได้ตาราง

ตาราง 9.5

	คะแนนการอ่านสำหรับปริมาณยา					
subject	ศูนย์	พอประมาณ	มาก	ผลรวม (X_{subj})	X^2_{subj}	$\overline{X}_{subj}$
1	34	59	30	123	15,129	41
2	24	56	21	101	10,201	33.67
3	29	65	24	118	13,924	39.33
4	40	45	45	130	16,900	43.33
5	50	78	39	167	27,889	55.67
6	35	74	41	150	22,500	50.00
7	30	62	26	118	13,924	39.33
8	25	34	30	89	7,921	29.67
9	45	77	62	184	33,856	61.33
10	15	57	26	98	9,604	32.67
11	45	80	36	161	25,921	53.67
12	24	64	32	120	14,400	40.00
	$\sum X$ = 396	751	412	$\sum X_{subj}$ 1559	$\sum X^2_{subj}$ 212,169	
	$\sum X^2$ = 14,274	30,150	13,879	$\sum X^2_{total}$ = 58303		
	n = 12	12	12			
	$\overline{X}$ = 33	62.58	34.33	$\overline{\overline{X}}$ = 43.30		

SOURCE	SS	df	s^2	F
Subject	2976.46	11	270.59	-
Between groups	6699.23	2	3349.61	82.87
Residual	889.3	22	40.42	
Total	10564.99	35		

ซึ่งการคำนวณค่าต่าง ๆในตารางอาจคำนวิณได้ดังนี้

$$SS_{total} = \sum^{all\ scores}(X-\bar{\bar{X}})^2 = (34-43.3)^2+(24-43.3)^2+(29-43.3)^2+(40-43.3)^2+(50-43.3)^2+(35-43.3)^2+(30-43.3)^2+(25-43.3)^2+(45-43.3)^2+(15-43.3)^2+(45-43.3)^2+(24-43.3)^2+(59-43.3)^2+(56-43.3)^2+(65-43.3)^2+(45-43.3)^2+(78-43.3)^2+(74-43.3)^2+(62-43.3)^2+(34-43.3)^2+(77-43.3)^2+(57-43.3)^2+(80-43.3)^2+(64-43.3)^2+(30-43.3)^2+(21-43.3)^2+(24-43.3)^2+(45-43.3)^2+(39-43.3)^2+(41-43.3)^2+(26-43.3)^2+(30-43.3)^2+(62-43.3)^2+(26-43.3)^2+(36-43.3)^2+(32-43.3)^2$$

$$= 86.49 + 372.49 + 204.49 + 10.89 + 44.89 + 68.89 + 176.89 + 334.89 + 2.89 + 800.89 + 2.89 + 372.49 + 246.49 + 161.29 + 470.89 + 2.89 + 1204.09 + 942.49 + 349.69 + 86.49 + 1135.69 + 187.9 + 1346.89 + 428.49 + 176.9 + 497.29 + 372.49 + 2.89 + 18.49 + 5.29 + 299.29 + 176.89 + 349.69 + 299.29 + 53.29 +127.69$$

$$= 10564.99$$

$$SS_{bet} = \sum^{k}(X_i-\bar{\bar{X}})^2 = 12(33-43.3)^2+12(62.58-43.3)^2+12(34.33-43.3)^2$$

$$= 1273.08+4460.62+965.53$$

$$= 6699.23$$

$SS_{subj} = \sum k(X_{subj} - \overline{\overline{X}})^2$

$= 3(41-43.3)^2 + 3(33.67-43.3)^2 + 3(39.33-43.3)^2 + 3(43.33-43.3)^2$

$+3(55.67-43.3)^2 +3(50-43.3)^2 +3(39.33-43.3)^2 +3(29.67-43.3)^2 +3(61.33-43.3)^2$

$+ 3(32.67-43.3)^2 +3(53.67-43.3)^2 + 3(40.00-43.3)^2$

= 15.87+ 92.74 +47.28+0.0027+459.05+134.67+338.99+

557.33+ 975.24+338.99+322.61+32.67

= 2976.46

$SS_{residual} = SS_{total} - SS_{bet} - SS_{subj}$

= 10564.99- 6699.23-2976.46

=889.3

หมายเหตุ

$\sum X^2$ (Z)=1156+576+841+1600+2500+1225+900+625+2025+225+2025+576

= 14,274

$\sum X^2$(M)=3481 + 3135 + 4225 + 2025 + 6084 + 5476 + 3844 + 1156 + 5929 +

3249 + 6400 + 4096

= 30,150

$\sum X^2$(H)= 900+441+576+2025+1521+676+900+3844+676+1296+1024

= 13,879

สังเกตว่า F = 40.42 จากตาราง C ในภาคผนวก เราหาค่าวิกฤดของ F ซึ่งจะได้ค่าวิกฤตของ F ที่องศาเสรีเป็น 2 และ 22 ที่ α = .05 คือ 3.44 ถ้าสมมุติฐานว่างเป็นจริงเพียง5% ของจำนวนครั้งที่เราจะได้รับค่า F ที่เท่ากับหรือมากกว่าค่านี้ เพราะฉะนั้น เรา

ปฏิเสธ H_0 และสรุปว่าความแตกต่างของขนาดของยาชนิดใหม่มีผลแตกต่างกันกับเด็กที่ว่องไวเกินปกติ กับการวิเคราะห์ความแปรปรวนทางเดียวที่ใช้การชักตัวอย่างแบบอิสระ ในตอนนี้เราต้องใช้การทดสอบ post hock เพื่อตรวจสอบว่ากลุ่มไหนแตกต่างจากกลุ่มอื่น

การทดสอบ HSD ของTukey อาจใช้กับ ANOVA สำหรับการวัดซ้ำ อย่างไรก็ดาม แทนที่จะใช้ s_w^2 เราใช้ s_{resid}^2 คำนวณ HSD

ค่าวิกฤต HSD สำหรับ การทดสอบ HSD ของ Turky ด้วยการวัดซ้ำ

$$\text{HSD} = q\sqrt{\frac{s_{resid}^2}{n}} \qquad \text{..........(9.18)}$$

ในตัวอย่างของเรา s_{resid}^2 = 40.42 และ n =12 ในการคำนวณ the Studentized range statistics ดูตาราง F ในภาคผนวก และหาค่า q สำหรับ df_{resid} =11และ k=3สำหรับ α =.05 ค่าของ q คือ 3.82 จะได้ค่าวิกฤต HSD

$$\text{HSD} = 3.82\sqrt{\frac{40.42}{12}} \approx 7.01$$

ค่านี้เป็นความแตกต่างที่น้อยที่สุดที่จำเป็นระหว่างค่าเฉลี่ยของสองกลุ่มใด ๆที่จะสรุปว่ามันมีความแตกต่างอย่างมีนัยสำคัญจากอีกกลุ่มหนึ่ง ถ้าเราพิจารณาตาราง 9.5 ปริมาณยาพอประมาณจะมีผลสูงกว่าอย่างมีนัยสำคัญในคะแนนการอ่าน ($\overline{X}$ = 62.58) สำหรับเด็กที่มีความว่องไวเกินปกติกว่าขนาดของยาเป็นศูนย์ ($\overline{X}$ = 33) และขนาดของยามาก ($\overline{X}$ = 34.33) ซึ่งไม่แตกต่างอย่างมีนัยสำคัญ

สรุป

การวิเคราะห์ความแปรปรวนทางเดียวอาจพิจารณาว่าเป็นการขยายของการทดสอบ t กับปัญหาที่เกี่ยวข้องกับกลุ่มตัวอย่างมากกว่าสองกลุ่ม หรือในทางกลับกัน การทดสอบ t อาจคิดเสมือนเป็นการใช้กรณีเฉพาะของการวิเคราะห์ความแปรปรวนทางเดียว

การวิเคราะห์ความแปรปรวนจะต้องหาผลบวกของกำลังสอง(SS) เสมือนหน่วยพื้นฐานของการวัด ในการวิเคราะห์ความแปรปรวนทางเดียวสำหรับการชักกลุ่มแบบอิสระ ผลบวกทั้งหมดของกำลังสอง (SS_{total}) สำหรับทุกคะแนน(กับกลุ่มทุกกลุ่ม) ถูกแบ่ง

กั้นเป็นสองส่วน : SS_{within} เป็นการวัดความเบี่ยงเบนภายในกลุ่มของคะแนนแต่ละคะแนน และ $SS_{between}$ เป็นการวัดความเบี่ยงเบนค่าเฉลี่ยของกลุ่มตัวอย่างระหว่างกลุ่ม เพื่อให้ได้ค่าการประมาณความแปรปรวน เราจะหาร SS ด้วยองศาเสรี เมื่อ H_0 เป็นจริง ($\mu_A = \mu_B = \mu_C$) ดังนั้น เมื่อ H_0 เป็นจริง s_w^2 และ s_{bet}^2 ควรเท่ากันภายในช่วงจำกัดของความคลาดเคลื่อนจากการสุ่ม เมื่อ H_0 เป็นเท็จ ความแปรปรวนภายในกลุ่มจะไม่เป็นผล (เพราะว่าทุกสมาชิกภายในกลุ่มยังคงทดสอบยังคงเหมือนเดิม(identically)) แต่ความแปรปรวนระหว่างกลุ่มในตอนนี้จะสะท้อนความแปรปรวนที่มีอยู่เดิม(inherent)บวกกับผลความแตกต่างในการปฏิบัติ) ผลการปฏิบัติมียิ่งมาก จะทำให้ค่าของ s_{bet}^2 ยิ่งมาก

ความแปรปรวนตรวจสอบจากอัตราส่วนที่เรียกว่าอัตราส่วน F (เพื่อเป็นเกียรติแก่ Fisher) ของสองค่าประมาณความแปรปรวนที่เป็นอิสระต่อกัน ถ้า H_0 เป็นจริง $F= \frac{s_{bet}^2}{s_w^2}$ จะมีค่าประมาณunit ถ้า H_0 เป็นเท็จ F จะมากกว่า unit (ผลของการปฏิบัติสะท้อนใน s_{bet}^2) มีทฤษฎีของการแจกแจง F ที่ขึ้นกับองศาเสรีสำหรับทั้ง s_{bet}^2 และ s_w^2 เช่นเดียวกับกรณีของ t, ANOVA ขึ้นอยู่กับสมมุติฐานของการชักตัวอย่างของการสุ่ม การแจกแจงปกติของประชากรของคะแนน และความเป็นเอกพันธ์ของความแปรปรวน

สมมุติฐานแย้ง(H_A) กำหนดเพียงว่าค่าเฉลี่ยของกลุ่มหนึ่งไม่เท่ากับบางกลุ่ม เมื่อได้นัยสำคัญ F จะต้องทดสอบ post hoc ในขั้นต่อไป เพื่อตรวจสอบว่ากลุ่มใดมีความแตกต่างกันระหว่างกลุ่มต่างๆ ซึ่งเราอาจตรวจสอบ post hoc ด้วยการทดสอบ HSD ของ Turkey ซึ่งจะตรวจสอบทุกคู่ระหว่างค่าเฉลี่ยของกลุ่มที่ปฏิบัติ

ในการวิเคราะห์ความแปรปรวนสำหรับการวัดซ้ำ s_w^2 แบ่งเป็นสองส่วนคือ s_{subj}^2 คือการวัดของความเบี่ยงเบนในคะแนนความแตกต่างของแต่ละตัว (individual) และ $s_{residual}^2$ คือการวัดความเบี่ยงเบนของความคลาดเคลื่อน อัตราส่วน F ใช้สำหรับ ANOVA นี้คือ s_{bet}^2 / s_{resid}^2

ข้อสังเกตทางคณิตศาสตร์

ข้อสังเกต 1. การแบ่งกั้นของผลบวกของกำลังสองและองศาเสรีในความแปรปรวนทางเดียว (อ้างอิง 3.)

$$(X - \bar{\bar{X}}) = (X - \bar{X}) + (\bar{X} - \bar{\bar{X}})$$

ยกกำลังสองทั้งสองด้าน เราจะได้

$$(X - \overline{\overline{X}})^2 = (X - \overline{X})^2 + 2(X - \overline{X})(\overline{X} - \overline{\overline{X}}) + (\overline{X} - \overline{\overline{X}})^2$$

หาผลบวกภายในแต่ละกลุ่ม

$$\sum^{n_i}(X - \overline{\overline{X}})^2 = \sum^{n_i}(X - \overline{X})^2 + 2(\overline{X} - \overline{\overline{X}})\sum^{n_i}(X - \overline{X}) + n_i(\overline{X} - \overline{\overline{X}})^2$$

เพราะว่า $\sum(X - \overline{X}) = 0$ พจน์กลางจะเป็นศูนย์ดังนั้นผลบวกบน k กลุ่ม จะได้

$$\sum^{k}\left[\sum^{n}(X - \overline{\overline{X}})^2\right] = \sum^{k}\left[\sum^{n_i}(X - \overline{X})^2\right] + \sum^{k} n(\overline{X} - \overline{\overline{X}})^2$$

หรือ

$$SS_{total} = SS_w + SS_{bet}$$

นับจำนวนองศาเสรีสำหรับแต่ละพจน์ เราจะได้

$$\sum^{k} n - 1 = \sum^{k}(n - 1) + (k-1)$$

หรือ

$$df_{total} = df_w + df_{bet}$$

พิสูจน์ประพจน์จาก

$$\sum^{k} n - 1 = \sum^{k} n - k + (k-1) = \sum^{k} n - \sum^{k} 1 + (k-1) = \sum^{k}(n-1) + (k-1)$$

แบบฝึกหัด 9

1. ในการวิเคราะห์ความแปรปรวนทางเดียว จำนวนเท่ากันของสมาชิก ถูกกระจายระหว่างกลุ่มต่าง ๆ

   1.1 ถ้า df_{bet} = 3 และ df_w = 84 จะมีกลุ่มทั้งหมดกี่กลุ่ม มีสมาชิก กี่สมาชิก ในแต่ละกลุ่ม ค่าของ df_{total} เป็นเท่าไร

   1.2 ถ้า s^2_{bet} = 170 และ s^2_w = 50 การปฏิบัติมีนัยสำคัญหรือไม่ ถ้า α =.05, ถ้า α =.01 การปฏิบัติมีนัยสำคัญหรือไม่

จงใช้ตาราง ตอบคำถามข้อ2 และ 3

มีสามสมาชิก ในแต่ละการปฏิบัติของสามการปฏิบัติ

		คะแนน	
กลุ่ม 1	2	6	3
กลุ่ม 2	7	8	6
กลุ่ม 3	10	12	8

2. จงใช้คะแนนในตารางต่อไปนี้ อธิบายดังตัวอย่าง

   2.1 ความแปรปรวนที่มีอยู่ประจำตัวคืออะไร

   2.2 ผลความแตกต่างของการปฏิบัติ

3. จงใช้สูตรคะแนนความเบี่ยงเบนสำหรับปัญหานี้สำหรับข้อมูลในตาราง

   3.1 หา SS_w , SS_{bet} และ SS_{total}

   3.2 SS_{total} เท่ากับ SS_w บวกกับ SS_{bet} หรือไม่

   3.3 จงหา df_w , df_{bet} และ df_{total}

   3.4 จงหา s^2_w และ s^2_{bet}

   3.5 จงนำเสนอผลการพัฒนาในตารางสรุปการวิเคราะห์ความแปรปรวน

   3.6 จงทดสอบสมมุติฐานว่า $\mu_1 = \mu_2 = \mu_3$ ที่ α = .05

4. ทำข้อ 3.ซ้ำโดยใช้ สูตรคะแนนดิบ

5. คุณต้องการตรวจสอบการควบคุมอาหารสามวิธีคือ ควบคุมอาหารให้มีไขมันต่ำ ควบคุมอาหารให้มีไขมันต่ำและเดินสี่กิโลเมตรในแต่ละวัน และควบคุมอาหารให้มีไขมันต่ำและวิ่งเหยาะสี่กิโลเมตรในแต่ละวัน คุณสุ่ม6 คนใส่ในแต่ละกลุ่ม บันทึกคะแนนดังตาราง แสดงการลดลงของคลอเรสเตอรอล ซึ่งตรวจสอบหลังจากสองเดือน

ตาราง

อาหารไขมันต่ำ	อาหารไขมันต่ำและเดิน	อาหารไขมันต่ำและวิ่งเหยาะ ๆ
11	15	20
8	11	14
14	18	14
13	17	16
9	12	18
6	18	20

5.1 จงเขียน H_0

5.2 จงทดสอบ F เมื่อกำหนด α =.05 และแสดงผลในตาราง

5.3 จงประยุกต์การทดสอบ HSD ของ Turkey

5.4 จงสรุปผล