

วิธีการเปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่หลังการวิเคราะห์ความแปรปรวน

Post-hoc Multiple Comparison Tests

ไพฑูรย์ สุขศรีงาม¹

Paitool Suksringarm¹

สารสังเขป

ในการทดสอบความแตกต่างค่าเฉลี่ยประชากรตั้งแต่ 3 ประชากรขึ้นไป โดยใช้การวิเคราะห์ความแปรปรวน เมื่อผลการทดสอบมีนัยสำคัญทางสถิติที่ระดับแอลฟา (α) ที่กำหนดแล้ว นักวิจัยต้องทำการทดสอบเพื่อให้ทราบว่าค่าเฉลี่ยของประชากรใดมีค่าแตกต่างจากค่าเฉลี่ยของประชากรที่เหลือ ซึ่งเรียกวิธีการนี้ว่าการเปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่ ทั้งนี้ต้องเลือกวิธีการเปรียบเทียบที่เหมาะสม มีการควบคุมค่าความผิดพลาดประเภทที่ 1 รวมให้มีค่าเท่ากับค่าแอลฟาที่กำหนดไว้ และมีค่าอำนาจในการทดสอบที่เหมาะสมด้วย ในบทความนี้ได้นำเสนอวิธีการต่าง ๆ ในการเปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่ เพื่อเป็นประโยชน์ในการเลือกใช้ให้เหมาะสมต่อไป

คำสำคัญ : การเปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่ ; ความผิดพลาดประเภทที่ 1 รวม ; ค่าอำนาจการทดสอบ

SUMMARY

In testing the difference of three or more population means by using the analysis of variance when the null hypothesis is rejected at a predetermined alpha level, called the test result is significant at the (α) level, the researcher has to test which population mean is different from the rest. This method is called the post-hoc multiple comparison tests, which can help the researcher to make conclusion about the greatest population mean. The researcher must select the appropriate method which can control the overall type I alpha rate equal to the predetermined alpha level and also reserve the power of the test. This article presents various methods of the post-hoc multiple comparison tests for the researcher to use appropriately in reporting the research results.

Keywords : Post-hoc Multiple Comparison Tests ; Overall Type I Error Rate ; Power of the Test.

¹ Ph.D.(Science Education) รองศาสตราจารย์ คณะสิ่งแวดล้อมและทรัพยากรศาสตร์ มหาวิทยาลัยมหาสารคาม
รับต้นฉบับ 1 มีนาคม 2557 รับตีพิมพ์ 1 เมษายน 2557



บทนำ

ในการทดสอบความแตกต่างค่าเฉลี่ยของประชากรจำนวนตั้งแต่ 3 ประชากรขึ้นไป โดยใช้สถิติทดสอบ F (F-test) สำหรับการวิเคราะห์ความแปรปรวน เป็นการทดสอบความแตกต่างค่าเฉลี่ยทุกค่าพร้อม ๆ กัน โดยมีการตั้งสมมติฐานว่างครั้งเดียว ($H_0 : \mu_1 \dots \mu_k$) บางครั้งเรียกสมมติฐานแบบนี้ว่า Complete หรือ Omnibus null hypothesis (Howell. 2007) ถ้าผลการทดสอบมีนัยสำคัญทางสถิติ หรือปฏิเสธสมมติฐานว่าง (Null hypothesis : H_0) แสดงว่าจะต้องมีความแตกต่างระหว่างค่าเฉลี่ยดังกล่าวเกิดขึ้น และวิธีการวิเคราะห์ความแปรปรวนดังกล่าวสามารถควบคุมอัตราความผิดพลาดของการตัดสินใจประเภทที่ 1 (Type I error rate หรือ Overall error rate) ให้คงที่เท่ากับค่าแอลฟาที่กำหนดไว้ (Pre-established α level) บางครั้งเรียกวิธีการทดสอบนี้ว่า Omnibus test (Hinkle, Wiersma and Jurs. 1994) หรือ Omnibus F-test (Sheskin. 2000) วิธีการทดสอบความแตกต่างค่าเฉลี่ยรายคู่ภายหลัง การวิเคราะห์ความแปรปรวนมีนัยสำคัญทางสถิติแล้ว เรียก Post-hoc multiple comparison test หรือ Pairwise comparisons (Ott. 1993 ; Hinkle Wiersma and Jurs. 1994 ; Field. 2009 ; Mayers, Well and Lorch. 2010) วิธีการทางสถิติสำหรับการเปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่นี้มีหลายวิธีการ ซึ่งทุกวิธีการที่ดีจะต้องควบคุมให้มีโอกาสการตัดสินใจประเภทที่ 1 (Overall α rate) ผิดพลาดมีค่าเท่ากับค่าแอลฟาที่กำหนดไว้ใน การวิเคราะห์ความแปรปรวน

อัตราความผิดพลาดประเภทที่ 1 (Type I Error Rate)

ในการทดสอบความแตกต่างค่าเฉลี่ยรายคู่ หลังการทดสอบสมมติฐานโดยใช้การวิเคราะห์ความแปรปรวนมีนัยสำคัญทางสถิติ วิธีการที่ใช้จะต้องสามารถควบคุมอัตราความผิดพลาดประเภทที่ 1 (Type I error rate) หรืออัตราความผิดพลาดรวม (Overall error rate) ให้มีค่าคงที่เท่ากับค่าแอลฟาหรือระดับนัยสำคัญที่กำหนดไว้ใน การวิเคราะห์ความแปรปรวน โดยทั่วไป **ความผิดพลาดประเภทที่ 1** นิยมแบ่งออกเป็น 2 แบบ **ความผิดพลาดแบบที่ 1** เรียก Comparison wise error rate (Hinkle, Wiersma and Jurs. 1994) Error rate per comparison

(PC) (Howell, 2007) หรือ Error rate per contrast (EC) (Mayers, Well and Lorch. 2010) เป็นอัตราความผิดพลาดประเภทที่ 1 ของการเปรียบเทียบความแตกต่างค่าเฉลี่ยแต่ละคู่ เช่น ในการทดสอบความแตกต่างค่าเฉลี่ย 2 ประชากร โดยใช้สถิติทดสอบ t และมีการปฏิเสธสมมติฐานว่าง (H_0) ที่ $\alpha = .05$ ดังนั้นที่มีอัตราความผิดพลาดของการเปรียบเทียบครั้งนี้เท่ากับ .05 **ความผิดพลาดแบบที่ 2** เรียก Familywise error rate (FWE หรือ FW) (Mayers, Well and Lorch. 2010 ; Howell. 2007) หรือ Experiment wise error rate (α E) (Hinkle, Wiersma and Jurs., 1994 and Freud and Wilson. 2010) ซึ่งเป็นโอกาสของการทำความผิดพลาดประเภทที่ 1 อย่างน้อยหนึ่งครั้งของชุดการเปรียบเทียบ หรือเป็นโอกาสของชุดการเปรียบเทียบ (Set or Family of contrasts) จำนวน 1 ชุด จะมีการตัดสินใจผิดพลาดประเภทที่ 1 เกิดขึ้น ซึ่งความผิดพลาดแบบนี้เกิดขึ้นในการเปรียบเทียบความแตกต่างค่าเฉลี่ยจำนวนหลายค่าหรือหลายกลุ่มค่าเฉลี่ย แล้วได้ข้อสรุปของการเปรียบเทียบดังกล่าว จำนวน 1 ชุด (Set หรือ Family) เช่น การเปรียบเทียบ $\mu_1 = \mu_2$, $\mu_2 = \mu_3$, $\mu_3 = \mu_4$ จัดเป็น 1 ชุดของการเปรียบเทียบ ซึ่งโอกาสของชุดข้อสรุปการเปรียบเทียบ ดังกล่าว จะมีอย่างน้อยที่เกิดโอกาสทำความผิดพลาดประเภทที่ 1 ความผิดพลาดนี้เรียก Familywise error rate (FWE)

อย่างไรก็ตามนักสถิติศาสตร์บางท่านได้ใช้คำว่า Experiment alpha level แทนโอกาสการทำความผิดพลาดประเภทที่ 1 รวม ซึ่งเกิดจากการสะสมความผิดพลาดจากการทดสอบสมมติฐานแต่ละครั้งแยกกัน ซึ่งค่ารวมดังกล่าว โดยทั่วไปมีค่ามากกว่าค่าแอลฟาที่กำหนดไว้จากการทดสอบแต่ละครั้ง (Gravetter and Wallnau. 2007)

ถ้ากำหนดให้ α แทนอัตราความผิดพลาดของการทดสอบความแตกต่างค่าเฉลี่ยรายคู่ 1 ครั้ง ดังนั้นอัตราความผิดพลาดของการทดสอบแต่ละครั้ง (PC) มีค่า $\alpha = \alpha$ และถ้ามีการทดสอบดังกล่าวจำนวนหลายครั้ง แต่ละครั้งมีความเป็นอิสระต่อกัน (Independent or orthogonal contrasts) เช่น ทดสอบจำนวน k ครั้ง สามารถคำนวณ FWE ได้ดังนี้ (Mayers, Well and Lorch. 2010)

$FWE = P$ (มีความผิดพลาดประเภทที่ 1 เกิดขึ้นจากชุดการทดสอบ)

$= 1 - P$ (ไม่มีความผิดพลาดประเภทที่ 1 จากชุดการทดสอบ)

$= 1 - P$ (ไม่มีความผิดพลาดประเภทที่ 1 ในการทดสอบหนึ่งครั้ง) k

ดังนั้น โอกาสของการตัดสินใจผิดพลาดประเภทที่ 1 ของการทดสอบหนึ่งครั้งก็คือ Comparison wise error rate (α') หรือ Error rate per contrast (EC) นั้นเอง โดยทั่วไปสูตรคำนวณ FWE มีดังนี้ (Sheskin. 2000)

Familywise error rate (FWE) = $1 - (1 - EC)^k$ หรือ Experiment wise error rate (α_E) = $1 - (1 - \alpha')^k$ โดย k เท่ากับ จำนวนครั้งของการเปรียบเทียบ และ α' เป็นอัตราความผิดพลาดของการทดสอบแต่ละครั้ง

ถ้าการเปรียบเทียบทุกคู่ไม่มีความเป็นอิสระต่อกัน เฉพาะค่า Error rate comparison (PC) จะมีค่าคงที่ไม่เปลี่ยนแปลง แต่ค่า FWE มีค่าเปลี่ยนแปลงเกิดขึ้น โดยทั่วไปสามารถกำหนดค่า FWE ไว้ดังนี้ $PC \leq FWE \leq k \alpha$ (Howell. 2007)

สมมติฐานว่าการทดสอบความแตกต่างเฉลี่ยรายคู่จำนวน 10 ครั้ง แต่ละครั้งใช้ Comparison wise error rate เท่ากับ α เมื่อแทนค่าในสูตรดังกล่าวจะได้ผลการคำนวณดังตารางข้างล่าง (Ott. 1993)

จำนวนครั้งของการทดสอบความแตกต่างค่าเฉลี่ยคู่	โอกาสของการเกิด Type I error rate		
	.10	.05	.01
1	.100	.050	.010
2	.190	.092	.020
3	.271	.143	.030
4	.344	.185	.039
5	.410	.226	.049
:	:	:	:
:	:	:	:
10	.651	.401	.096

จากตารางพบว่า ถ้าพิจารณาการทดสอบทั้ง 10 ครั้ง จะได้ค่า Familywise error rate (FWE) เท่ากับ .651 .401 หรือ .096 ถ้ากำหนด Comparison wise error rate (α) เท่ากับ

.10 .05 หรือ .01 ตามลำดับ นั่นก็คือ ถ้าค่าเฉลี่ยประชากรมีค่าเท่ากัน โอกาสของการตัดสินใจ ผิดพลาดอย่างน้อย 1 ครั้ง หรือมากกว่าจากการทดสอบรวม 10 ครั้ง มีค่าเท่ากับ .651 .401 หรือ .096 ซึ่งจะเห็นได้ว่าค่า FWE มีค่ามากกว่าค่า α (หรือ Comparison wise error rate หรือ Error rate per contrast : EC) ดังนั้น วิธีการเปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่ที่เหมาะสมจะต้องควบคุมโอกาสของการตัดสินใจผิดพลาดประเภทที่ 1 (FWE) ของการทดสอบทั้งชุดมากกว่าการควบคุมความผิดพลาดของการทดสอบแต่ละครั้ง (EC) นั้นเอง

อย่างไรก็ตามในการทดสอบความแตกต่างค่าเฉลี่ยรายคู่สืบเนื่องจากการวิเคราะห์ความแปรปรวน เป็นการทดสอบที่ไม่เป็นอิสระซึ่ง FWE อาจจะมีค่าน้อยกว่า $1 - (1 - \alpha)^k$ จึงมีการประมาณค่า FWE โดยใช้สูตร (Hinkle Wiersma and Jurs. 1994 and Sheskin. 2000)

$$\alpha' = k(\alpha) \dots\dots\dots(1)$$

สูตรนี้สามารถคำนวณหา α' ที่ใช้ในการทดสอบความแตกต่างค่าเฉลี่ยแต่ละครั้ง (เรียก Comparison wise error rate) ให้มีค่าเท่ากับค่าแอลฟา ที่กำหนดไว้ในการวิเคราะห์ความแปรปรวน (Pre- established level) ได้โดยการนำจำนวนครั้งของการทดสอบ (k) ไปหาร FWE เช่นถ้า FWE มีค่าเท่ากับ .05 ดังนั้น Comparison-wise error rate (α) ของการทดสอบแต่ละครั้งจะมีค่าเท่ากับ $.05/10 = .005$ ดังนั้น จึงคำนวณหาค่า Comparison wise error rate จากสูตร

$$\alpha = \frac{\alpha'E}{k} \dots\dots\dots(2)$$

ซึ่งสูตรนี้เรียก Bonferroni correction (Field. : 2009) โปรดสังเกตว่ากระบวนการควบคุม FWE ดังกล่าวมาแล้ว ก่อนข้างเป็นการอนุรักษ์อย่างมาก (Extremely conservative) ซึ่งอาจมีผลทำให้การทดสอบความแตกต่างค่าเฉลี่ยรายคู่ไม่มีนัยสำคัญทางสถิติเกิดขึ้น แม้ว่าผลการทดสอบ F-test จากการวิเคราะห์ความแปรปรวนมีนัยสำคัญทางสถิติก็ตาม นอกจากนี้ การปรับให้ค่าความผิดพลาดประเภทที่ 1 ลดลงจะมีผลทำให้ค่าความผิดพลาดประเภทที่ 2 (Type II error) เพิ่มขึ้นหรือค่าอำนาจการทดสอบ (Power of the test) มีค่าลดลงด้วย

เนื่องจากมีวิธีการทางสถิติหลายวิธีที่ใช้ในการทดสอบความแตกต่างค่าเฉลี่ยรายคู่ นักวิจัยจึงควรใช้เกณฑ์เหล่านี้ใน



การเลือกวิธีการที่เหมาะสม (Field. 2009)

1. วิธีการทดสอบสามารถควบคุม FWE ได้หรือไม่
2. วิธีการทดสอบสามารถควบคุมความผิดพลาดประเภทที่ 2 (Type II error) หรืออำนาจการทดสอบได้หรือไม่
3. วิธีการนี้เชื่อถือได้หรือไม่ ถ้าข้อมูลไม่สอดคล้องกับข้อตกลงเบื้องต้นบางประการ (Assumptions) ของการวิเคราะห์ความแปรปรวนจำนวนคู่หรือจำนวนครั้ง ของการเปรียบเทียบความ

$$\text{ความแตกต่างค่าเฉลี่ยหาได้จากสูตร } \frac{k - (k - 1)}{2}$$

(Hinlke Wiersma and Jurs. 1994 and Sheskin. 2000)

เมื่อ K แทน จำนวนค่าเฉลี่ย

ชนิดของการเปรียบเทียบความแตกต่างค่าเฉลี่ย (Types of Comparisons)

การเปรียบเทียบความแตกต่างค่าเฉลี่ยประชากรโดยใช้ omnibus F-test นักวิจัยสามารถเลือกการเปรียบเทียบความแตกต่างค่าเฉลี่ยประชากรก่อนหรือหลังทำการวิเคราะห์ด้วย Omnibus F-test โดยการเปรียบเทียบค่าเฉลี่ยดังกล่าวมี 2 แบบ (Sheskin. 200 ; Gravetter and Wallnau. 2007 and Howell. 2007)

1. การเปรียบเทียบแบบวางแผนไว้ล่วงหน้า (Planned comparisons หรือ A priori comparisons) การเปรียบเทียบตามวิธีนี้นักวิจัยจะกำหนดการเปรียบเทียบค่าเฉลี่ยประชากรบางคู่ไว้ก่อนที่จะดำเนินการเก็บรวบรวมข้อมูล เมื่อเก็บข้อมูลแล้วนักวิจัยก็ทำการเปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่ตามที่กำหนดไว้ ต่อมาจึงทำการคำนวณ Omnibus F-test โดยไม่คำนึงว่าผลการทดสอบจะมีนัยสำคัญทางสถิติหรือไม่ และไม่จำเป็นต้องปรับค่า Familywise error rate (FWE) แต่บางท่านเสนอแนะให้ใช้จำนวนครั้งของการทดสอบหารค่าแอลฟาของ Omnibus F-test แล้วใช้เป็นเกณฑ์ในการเปรียบเทียบความแตกต่างค่าเฉลี่ยแต่ละคู่ (Gravetter and Wallnau. 2007). ในทางปฏิบัติการเปรียบเทียบแบบวางแผนไว้ล่วงหน้า แบ่งออกเป็นระดับย่อย คือ การเปรียบเทียบอย่างง่าย (Simple comparisons) กับการเปรียบเทียบเชิงซ้อน (Complex comparisons) โดยการเปรียบเทียบอย่างง่ายเป็นการเปรียบเทียบ

ความแตกต่างเฉลี่ยครั้งละ 1 คู่ เช่น $\bar{X}_1$ กับ $\bar{X}_2$ $\bar{X}_2$ กับ $\bar{X}_3$ ฯลฯ การเปรียบเทียบแบบนี้นิยมเรียก Pairwise comparisons

ส่วนการเปรียบเทียบเชิงซ้อนเป็นการเปรียบเทียบค่าเฉลี่ยโดยมีการรวมค่าเฉลี่ย 2 ค่าหรือมากกว่า เปรียบเทียบกับค่าเฉลี่ยหนึ่งค่าเฉลี่ยรวมกันของค่าเฉลี่ย 2 ค่า หรือมากกว่าอีกกลุ่ม เช่น $\bar{X}_1$ กับ $(\bar{X}_2 + \bar{X}_3)/2$; $(\bar{X}_1 + \bar{X}_2)/2$ กับ $(\bar{X}_3 + \bar{X}_4)/2$ เป็นต้น

2. การเปรียบเทียบแบบไม่มีการวางแผนล่วงหน้า (Unplanned comparisons หรือ Posthoc, multiple or A posteriori comparisons) การเปรียบเทียบนี้ไม่ว่าจะเป็นชนิดการเปรียบเทียบอย่างง่ายหรือการเปรียบเทียบเชิงซ้อน เป็นการเปรียบเทียบความแตกต่างค่าเฉลี่ยหลังจากเก็บข้อมูลแล้ว และส่วนมากเป็นการเปรียบเทียบค่าเฉลี่ยจำนวนมาก นักสถิติศาสตร์ส่วนมากแนะนำให้ทำการเปรียบเทียบค่าเฉลี่ยรายคู่หลังจาก Omnibus F-test มีนัยสำคัญทางสถิติแล้ว แต่ก็ยังมีนักสถิติศาสตร์บางท่านแนะนำให้ทำการเปรียบเทียบค่าเฉลี่ยได้โดยไม่คำนึงว่า Omnibus F-test จะมีนัยสำคัญทางสถิติหรือไม่ แต่ทุกท่านเสนอว่า จะต้องมีการปรับหรือควบคุม Familywise error rate (FWE) แต่ยังมีความเห็นแตกต่างกันว่า วิธีการเปรียบเทียบความแตกต่างค่าเฉลี่ยของใครดีที่สุดในการควบคุมหรือปรับ FWE

การเปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่ ไม่ว่าจะ เป็นชนิดอย่างง่ายหรือชนิดเชิงซ้อน เป็นการเปรียบเทียบโดยใช้ผลรวมเชิงเส้น (Linear combination) ของค่าเฉลี่ยประชากร เรียกการเปรียบเทียบแบบนี้ว่าการเปรียบเทียบเชิงเส้น (Linear contrasts) เช่นกรณีการเปรียบเทียบอย่างง่าย การเปรียบเทียบเชิงเส้น ได้แก่ การเปรียบเทียบค่าเฉลี่ย 2 ค่า เช่น $\bar{X}_1$ กับ $\bar{X}_2$ ส่วนการเปรียบเทียบเชิงซ้อน จะมีการรวมค่าเฉลี่ย 2 ค่า หรือมากกว่าเปรียบเทียบกับค่าเฉลี่ยหนึ่งค่าหรือการรวมค่าเฉลี่ยของอีกกลุ่มหนึ่ง เช่น $\bar{X}_1$ กับ $(\bar{X}_2 + \bar{X}_3)/2$ เป็นต้น

วิธีทางสถิติสำหรับการทดสอบความแตกต่างค่าเฉลี่ยรายคู่

วิธีการทางสถิติที่จะกล่าวต่อไปนี้เป็นเพียงบางวิธีเท่านั้น ซึ่งแต่ละวิธีสามารถควบคุมอัตราความผิดพลาดรวม (Overall error rate หรือ Familywise error rate) ให้มีค่าเท่ากับระดับนัยสำคัญที่กำหนดไว้จากการวิเคราะห์ความแปรปรวน (ค่า α)

1. วิธีการของ Tukey

วิธีการของ Tukey. (1953) มีชื่อเรียกเฉพาะว่า Honestly Significant Difference : HSD) ใช้ในกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากัน (Equal sample size) การควบคุม FWE ใช้การแจกแจง Q (Q distribution) โดย Q เรียก Studentized range statistic หรือ Studentized Q statistic สำหรับการแจกแจงค่าสถิติ (Sampling distribution) โดยสถิติ Q หาได้ดังนี้

$$Q = \frac{\bar{X}_a - \bar{X}_b}{\sqrt{\frac{MSW}{n}}} \dots\dots\dots(3)$$

โดย $\bar{X}_a$ และ $\bar{X}_b$ แทน ค่าเฉลี่ย MSW แทน Within group mean square บางครั้งก็เรียกว่ามีการใช้ Studentized range distribution โดยหาจากสูตร (Ott. 1993)

$$Q = \frac{\bar{X}_a - \bar{X}_b}{S_p \sqrt{\frac{1}{n}}} \dots\dots\dots(4)$$

SP แทน Pooled standard deviation

และ

$$HSD = Q \sqrt{\frac{MSW}{n}} \dots\dots\dots(5)$$

(Gravetter and Wallnau. 2007)

ในกรณีที่ขนาดกลุ่มตัวอย่างไม่เท่ากัน (Unequal sample size) ใช้วิธีการของ Tukey - Kramer (TK) method โดยมีสูตร ดังนี้ (Hinkle Wiersma and Jurs. 1994 and Howell. 2007)

$$Q = \frac{\bar{X}_{\max} - \bar{X}_{\min}}{\sqrt{MSW \left[\frac{1}{n_{\max}} + \frac{1}{n_{\min}} \right]}} \dots\dots\dots(6)$$

2. วิธีการของ Newman-Keuls

วิธีการของ Newman-Keuls บางครั้งเรียก Student-Newman-Keul (SNK) ปรับปรุงมาจากวิธีการของ Tukey โดยมีการหาสถิติ Studentized range statistic แต่มีการกำหนดค่าวิกฤตหลายค่า สำหรับการเปรียบเทียบกับค่าสถิติที่ได้จากการคำนวณแต่ละค่าขึ้นอยู่กับจำนวนขั้นที่แบ่งแยกค่าเฉลี่ยที่นำมาเปรียบเทียบกับกัน วิธีการนี้สามารถใช้ได้ทั้งกรณีกลุ่มตัวอย่างมีขนาดเท่ากันหรือไม่เท่ากัน ที่น่าสนใจก็คือ วิธีการ SNK ไม่มีทั้ง Experimentwise error rate และไม่มี Comparisonwise error rate แต่มีการกำหนดอัตราความผิดพลาด (Error rate) สำหรับค่าเฉลี่ยที่มีจำนวนขั้นที่เรียงลำดับเหมือนกับที่แบ่งแยก (Ott. 1993) เนื่องจากค่าวิกฤตมีค่าลดลง เมื่อมีจำนวนขั้นระหว่างค่าเฉลี่ยเปรียบเทียบมีค่าลดลง ทำให้วิธีการของ SNK มีความเป็นอนุรักษ์ลดลง จะทำให้ผลการเปรียบเทียบความแตกต่าง พบมากกว่าวิธีการของ Tukey ซึ่งมีการใช้ค่าวิกฤตที่มีขนาดใหญ่ที่สุดเพียงค่าเดียวในการเปรียบเทียบกับค่าสถิติจากขั้นต่าง ๆ

อย่างไรก็ดีการที่ SNK สามารถพบจำนวนคู่ของการเปรียบเทียบความแตกต่างค่าเฉลี่ยได้มากกว่า หรือ มีการอนุรักษ์น้อยกว่าวิธีการของ Tukey. (อ้างถึงใน Todman and Dugard. 2007) นับว่าเป็นวิธีการที่เหมาะสมมากกว่า ทั้งนี้เนื่องจาก FWE ของ SNK จะมีค่ามากกว่าเล็กน้อย จากค่าแอลฟาที่กำหนดในการวิเคราะห์ความแปรปรวน (Howell. 2007) แต่ก็ยังมีค่าน้อยกว่าการทดสอบโดยใช้ t-test ส่วนวิธีการของ Tukey. จะมีค่า FWE เท่ากับค่าแอลฟาที่กำหนดไว้ (Hinrle, Weirsmas and Jurs. 1994)

3. วิธีการของ Fisher

วิธีการของ Fisher. (1949) เรียก Least Significant Difference (LSD) สามารถใช้เปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่ไว้ล่วงหน้าก่อนทำการวิเคราะห์ความแปรปรวน (Planned comparisons) หรือ ภายหลังการวิเคราะห์ความแปรปรวน (Post - hoc comparisons) เฉพาะกรณีใช้ภายหลังการวิเคราะห์ความแปรปรวนแล้ว เป็นวิธีการที่ได้ปรับปรุงแล้ว จึงเรียกชื่อใหม่ว่า Fisher's Protected LSD หรือ Fisher's Protected t ซึ่งสามารถควบคุม FWE ให้มีค่าเท่ากับค่าแอลฟาที่กำหนดสำหรับ F-test ได้ (Ott. 1993 : Freud and Wilson. 2010) และใช้ t-test เปรียบเทียบความแตกต่างค่าเฉลี่ยแต่ละ



คู่โดยมีสูตรดังนี้

$$LSD = \frac{\alpha}{2} t_{df} \sqrt{MSM \left[\frac{1}{n_a} - \frac{1}{n_b} \right]} \dots\dots\dots(7)$$

กรณีกลุ่มตัวอย่างขนาดเท่ากัน

$$LSD = \frac{\alpha}{2} t_{df} \sqrt{\frac{2MSW}{n}} \dots\dots\dots(8)$$

จากการศึกษาพบว่า วิธีการของ Tukey's HSD มีความเป็นอนุรักษ์มากกว่าวิธีการของ Fisher's LSD ทำให้มีจำนวนคู่ของการเปรียบเทียบที่มีนัยสำคัญทางสถิติต่ำกว่า แม้ว่าทั้งสองวิธีการมี FWE เหมือนกันแต่ LSD มีขนาดของการตัดสินใจผิดพลาดแต่ละครั้ง (Comparisonwise error rate) ใหญ่กว่าของ Tukey's HSD นั่นเอง

อย่างไรก็ตาม Howell. (2007) ได้แนะนำให้ใช้วิธีการของ Fisher เฉพาะกรณีมีค่าเฉลี่ยไม่เกิน 3 ค่า เท่านั้น เพราะสามารถควบคุม FWE ให้เท่ากับค่าแอลฟาที่กำหนดไว้

4. วิธีการของ Scheffe'

วิธีการของ Scheffe' (1959) สามารถใช้เปรียบเทียบความแตกต่างค่าเฉลี่ยได้มากกว่าที่ละคู่ จึงมีการเปรียบเทียบได้อย่างซับซ้อน และมีความเป็นอนุรักษ์ค่อนข้างมาก (More conservative) เมื่อเปรียบเทียบกับวิธีการของคนอื่น เนื่องจากมีจำนวนคู่ของการเปรียบเทียบในคราวเดียวกับจำนวนมาก วิธีการนี้สามารถใช้กับการฝึกกลุ่มตัวอย่างมีขนาดเท่ากัน หรือไม่เท่ากันก็ได้ วิธีการของ Scheffe. สามารถควบคุม FWE ให้มีค่าคงที่เท่ากับค่าระดับนัยสำคัญทางสถิติ (หรือ α) ที่กำหนดไว้ในการวิเคราะห์ความแปรปรวน และมีความแกร่ง (Robustness) ต่อการแจกแจงข้อมูลประชากรไม่เป็นโค้งปกติ หรือมีความแปรปรวนประชากรไม่เท่ากัน (Wampole and Drew. 1990) อย่างไรก็ตามวิธีการของ Scheffe' อาจไม่พบความแตกต่างของค่าเฉลี่ยรายคู่ก็ได้

วิธีการของ Scheff' ใช้การแจกแจง F แทน Studentized range test และสามารถควบคุม FWE ได้อย่างคงที่ แต่มีอำนาจทดสอบน้อยกว่าวิธีการของคนอื่น ๆ (Howell. 2007)

5. วิธีการของ Duncan

วิธีการของ Duncan. (1955) มีชื่อเฉพาะว่า New

Multiple Range Test (NMRT) เป็นวิธีการเปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่ที่ใช้ Studentized range distribution โดยค่าอัตราผิดพลาดในการตัดสินใจจะไม่ใช้ทั้ง Familywise Error rate (FWE) และไม่ใช่ Comparisonwise error rate วิธีการนี้มีการเรียงลำดับค่าเฉลี่ยกลุ่มตัวอย่างจากน้อยที่สุดไปมากที่สุด แล้วได้กำหนดอัตราความผิดพลาดในการตัดสินใจ (Error rate) จากค่าระดับการป้องกัน (Protection level) ดังนี้ (Ott. 1993 ; Freud and Wilson. 2010)

$$\text{error rate} = 1-(1-\alpha)^{r-1} \dots\dots\dots(9)$$

โดย $(1-\alpha)^{r-1}$ เรียก Protection level ซึ่งมีค่าลดลงเมื่อค่า r มีค่ามากขึ้น

r แทนจำนวนขั้นแบ่งแยกค่าเฉลี่ย

จากสูตร แสดงให้เห็นถึงโอกาสของการปฏิเสธสมมติฐานว่าง ($H_0 : \mu_1 = \mu_2$) อย่างผิดพลาด เมื่อค่าเฉลี่ยทั้ง 2 ค่าอยู่ห่างกัน r ขั้น ดังนั้นโอกาสของการทำความผิดพลาดประเภทที่ 1 เมื่อมีค่าเฉลี่ยตัวอย่างเท่ากับ r ขั้น มีค่า $1-(1-\alpha)^{r-1}$ ถ้าค่าเฉลี่ย 2 ค่า อยู่เรียงติดกัน ($r=2$) จะได้ Protection level เท่ากับ $1-(1-\alpha) = \alpha$

และจากการพบว่า ค่า Protection level มีค่าลดลงเรื่อย ๆ สัมพันธ์กับจำนวนค่าเฉลี่ยที่จะนำมาเปรียบเทียบความแตกต่าง Duncan. จึงกำหนดให้ค่าแอลฟาเท่ากับปริมาณของ $1-(1-\alpha)^{r-1}$ หรือ Error rate เมื่อคู่ใด ๆ ของค่าเฉลี่ยที่จะเปรียบเทียบอยู่ห่างกันหรือแบ่งแยก r ขั้น ($r = 2, \dots, k$) การปรับเปลี่ยนตามแนวคิดนี้จะทำให้พบว่า จะไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติของค่าเฉลี่ยที่เรียงอยู่ติดกัน แต่จะพบเฉพาะค่าเฉลี่ยที่ถูกแบ่งแยกจากกันหลายขั้น (Milton. 1992)

วิธีการของ Duncan. ปกติใช้ในการฝึกกลุ่มตัวอย่างมีขนาดเท่ากัน ต่อมาได้มีการปรับปรุงให้สามารถใช้ได้กับการฝึกกลุ่มตัวอย่างมีขนาดไม่เท่ากันเรียก Duncan - Kramer method (1957)

ถ้าผลการวิเคราะห์โดยใช้วิธีการของ Duncan. เปรียบเทียบกับวิธีการของ Fisher's LSD, Tukey's HSD และ Newman-Keul's SNK จะพบว่าผลการเปรียบเทียบความแตกต่าง

ต่างเฉลี่ยรายคู่ตามวิธีของ Duncan. ได้ผลเหมือนกับวิธี LSD และ SNK มากกว่าวิธีการ HSD และมีอำนาจในการทดสอบมากกว่า (Ott. 1993)

6. วิธีการอื่น ๆ

6.1 ในกรณีที่ประชากรมีความแปรปรวนไม่เท่ากัน นิยมใช้วิธีการของ Howell. (1976) ซึ่งสามารถควบคุม FWE ได้อย่างเหมาะสม อย่างไรก็ตามบางสถานการณ์วิธีการนี้อาจทำให้ FWE มีค่ามากกว่าเล็กน้อยจากค่าแอลฟาที่กำหนดไว้ เช่น ถ้าประชากรมีความแปรปรวนเท่ากัน และกลุ่มตัวอย่างมีขนาดเล็กกว่า 50 ค่า FWE จะมีค่าเกิน .05 เล็กน้อยเช่นเป็น .07 (Myers, Well and Lorch. 2010) หรือถ้าความแปรปรวนประชากรไม่เท่ากัน และกลุ่มตัวอย่างมีขนาดน้อยกว่า 6 ค่า FWE ก็จะมีค่ามากกว่าค่าแอลฟาเล็กน้อย ที่น่าสนใจคือวิธีการของ Howell. (1976) มีอำนาจในการทดสอบสูงกว่าวิธีอื่น ๆ

โปรแกรมสำเร็จรูปคำนวณสถิติ (SPSS) ได้กำหนดวิธีการทดสอบความแตกต่างค่าเฉลี่ยดังกล่าวมาแล้วในกรณีประชากรมีความแปรปรวนต่างกันไว้ 4 วิธี คือ Tamhane's and Howell. โดยวิธีการของ Tamhane's T₂, มีความอนุรักษ์ค่อนข้างมาก ส่วนวิธี Dunnett's T₃, Dunnett' C และ และ C สามารถควบคุมความผิดพลาดประเภทที่ 1 ได้ดีมาก และวิธีการของ Howell มีอำนาจการทดสอบมากที่สุด (Fidell. 2007)

6.2 กรณีกลุ่มตัวอย่างมีขนาดไม่เท่ากัน แต่ขนาดแตกต่างไม่มากนัก นิยมใช้วิธีการของ Hochberg's GT₂ และ Gabrieli's procedure โดยวิธีของ Gabriel. มีอำนาจในการทดสอบสูง ถ้าขนาดกลุ่มตัวอย่างแตกต่างกันมาก แต่ความแปรปรวนประชากรเท่ากันใช้วิธีการของ Hochberg's procedure (Field. 2009)

6.3 กรณีกลุ่มตัวอย่างมีขนาดเท่ากัน นอกจากใช้วิธีการบางวิธีดังกล่าวมาแล้ว สามารถใช้วิธีการของ Ryan, Einot, Gabriel and Welsch Q procedure (REGWQ) ซึ่งสามารถควบคุม FWE ได้ดี และมีอำนาจการทดสอบสูง (Field. 2009) Howell. (2007) แนะนำให้เลือกใช้วิธีของ REGWQ แทนการใช้วิธีของ SNK และ Tukey's HSD เพราะเป็นวิธีการที่มีอำนาจการทดสอบมากที่สุด และสามารถควบคุม FWE เท่ากับค่าแอลฟาที่กำหนดไว้

สรุป

ในการทดสอบความแตกต่างค่าเฉลี่ยรายคู่ หลังการทดสอบสมมติฐานโดยใช้การวิเคราะห์ความแปรปรวนมีนัยสำคัญทางสถิติ นักวิจัยจะต้องเลือกวิธีการที่เหมาะสมในการควบคุม Familywise error rate (FWE) ให้มีค่าคงที่เท่ากับค่าแอลฟา หรือระดับนัยสำคัญทางสถิติที่กำหนดไว้ในกรณีวิเคราะห์ความแปรปรวน ซึ่งนักสถิติศาสตร์ได้วิเคราะห์และพบว่า มีปัจจัย 2 ประการที่มีผลกระทบต่อค่าความผิดพลาด (Overall error rate) คือ ความเป็นวิวิธพันธุ์ของความแปรปรวนประชากร (Heterogeneity of variance) หรือการมีความแปรปรวนประชากรไม่เท่ากัน และการมีขนาดกลุ่มตัวอย่างไม่เท่ากัน ดังนั้นจึงต้องหาวิธีการควบคุม FWE ดังกล่าว ในการเลือกวิธีการใด นักวิจัยจะต้องเลือกวิธีการที่ควบคุม FWE ได้อย่างเหมาะสม และวิธีการนั้นมีอำนาจในการทดสอบ (Power of the test) สูงด้วย

เอกสารอ้างอิง

- Field. A. (2009) **Discovering Statistics Using SPSS**. 3rd ed., California ; SAGE Publication Ltd.
- Freund, F.J., W.J. Wilson and D.L. Mohr. (2010) **Statistical Methods**. Singapore : Academic Press
- Gravetter, F. J. and L.B. Wallnau. (2007) **Statistics for the Behavioral Sciences**. 7th ed., Singapore : Thomson Wadsworth.
- Hinre, D.E. ; W. Wiersma and S.G. Jurs. (1994) **Applied Statistics for the Behavioral Sciences**. Boston : Houghton Mifflin Company
- Howell. D.C. (2007) **Statistical Methods for Psychology**. 6th ed., Singapore: Thomson. Wadsworth.
- Myers. J.L. ; A.D. Well and R.F. Lorch. (2010) **Research Design and Statistical Analysis**. 3rd ed., New York : Routledge.
- Milton, J.S. (1992) **Statistical Methods in the Biological and Health Sciences**. New Jersey : Mc Graw - Hill International Series.



Ott, L.D (1993) **An Introduction to Statistical Methods and Data Analysis**. 4th ed., Belmont : Duxbury Press.

Sheskin, D.J. (2000) **Handbook of Parametric and Nonparametric Statistical Procedures**. 2nd ed. New York : Chapman and Hall.

Todman, J. and P. Dugard. (2007) **Approaching Multivariate Analysis : An Introduction for Psychology**. New York : Psychology Press.

Wampold BE. and C.J., Drew. (1990) **Theory and Application of Statistics**. New York : Mc Graw - Hill Publishing Company.