

การเปรียบเทียบก่อนและหลังการทดสอบรวมเพื่อการวิจัย (Priori and Posteriori Comparisons for a Research Study)

รองศาสตราจารย์ ดร. สุวัฒน์ สุขมลสันต์

บทคัดย่อ

ผู้วิจัยจำนวนมากต้องการที่จะเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างภายหลังหรือก่อนการทดสอบรวม แต่มักมีปัญหาในการเลือกใช้วิธีการเปรียบเทียบที่เหมาะสมเนื่องจากมีวิธีการเปรียบเทียบจำนวนมากเพื่อใช้ในเงื่อนไขที่แตกต่างกัน ดังนั้นบทความนี้จึงนำเสนอการเลือกใช้การเปรียบเทียบก่อนและหลังการทดสอบรวมเพื่อการวิจัย และข้อเสนอแนะทั่วไปในการเลือกใช้วิธีต่างๆเพื่อให้เหมาะสมกับเงื่อนไขต่างๆของการวิจัย เพราะหากใช้วิธีที่ไม่เหมาะสมผลการวิจัยอาจแตกต่างกันได้

คำสำคัญ: สถิติในการทำวิจัย, วิธีวิจัย

Abstract

Many researchers want to compare means of different groups after or before an omnibus test but they often have problems in selecting an appropriate method since there are many comparing methods for different conditions. Therefore, this article presents how to select an appropriate method for comparing means of different groups before and after an omnibus test, and also gives some general suggestions for selecting an appropriate method to best suit certain testing conditions of the groups; otherwise, using an inappropriate method may give the researchers different results.

Key words: statistics in research, research methods

ความนำ

ปกติแล้วเมื่อนักวิจัยทำการศึกษาเปรียบเทียบผลการวิจัยที่มีกลุ่มพลวิจัย (subjects) หรือกลุ่มตัวอย่าง (samples) มากกว่า 2 กลุ่ม เช่น 3 กลุ่ม มักต้องการจะรู้ว่า 1) ผลการทดลองทั้ง 3 กลุ่มโดยเฉลี่ยแล้วแตกต่างกันอย่างมีนัยสำคัญหรือไม่ และ 2) หากว่าผลการวิจัยแตกต่างกันกลุ่มใดแตกต่างจากกลุ่มใด และกลุ่มใดไม่แตกต่างจากกลุ่มใดอย่างมีนัยสำคัญบ้างหรือไม่

ในกรณีที่ 1 นั้น การทดสอบความแตกต่างของค่าเฉลี่ยนิยมทำการทดสอบโดยสถิติการทดสอบรวม (Omnibus Test) ซึ่งหมายถึงการทดสอบว่าค่าเฉลี่ยทั้งหมดของกลุ่มตัวอย่างในการวิจัยแตกต่างกันอย่างมีนัยสำคัญหรือไม่ เช่น ANOVA (Analysis of Variance) หรือ ANCOVA (Analysis of Covariance) หรือที่นิยมเรียกกันทั่วไปว่า F-test ซึ่งมีวิธีแตกต่างกันแล้วแต่รูปแบบของการทดลอง (Experimental Design)

สำหรับกรณีที่ 2 เป็นการเปรียบเทียบค่าเฉลี่ยของผลการวิจัยรายย่อย ซึ่งอาจเรียกว่าการทดสอบย่อยและมีสถิติใช้ทดสอบจำนวนประมาณ 20 ชื่อ และทำได้ 2 กรณี (GraphPad, 2010; Karpinski, 2006) คือ

1. การเปรียบเทียบค่าเฉลี่ยก่อนการทดสอบรวม เรียกว่า การเปรียบเทียบก่อน (Prior Tests) หรือการเปรียบเทียบตามแผนการที่วางไว้ (Planned Comparison) หรือการเปรียบเทียบความต่าง (Contrast Comparison) ด้วยสถิติหลายอย่างเช่น Bonferroni Test, t-test และ Sidak Test เป็นต้น

2. การเปรียบเทียบค่าเฉลี่ยภายหลังการทดสอบรวม เรียกว่าการเปรียบเทียบภายหลัง (Posteriori Tests หรือ Post Hoc Tests) หรือการเปรียบเทียบหลากหลาย (Multiple Comparison) ด้วยสถิติหลายชนิดตามเงื่อนไขที่แตกต่างกัน เช่น

ก. เมื่อความแปรปรวนของคะแนนของกลุ่มที่ต้องการเปรียบเทียบไม่ต่างกัน ($\delta_1^2 = \delta_2^2$) อาจเลือกใช้สถิติ Bonferroni, Scheffe, Tukey หรือ Duncan เป็นต้น

ข. เมื่อความแปรปรวนของคะแนนของกลุ่มที่ต้องการเปรียบเทียบแตกต่างกันอย่างมีนัยสำคัญ ($\delta_1^2 \neq \delta_2^2$) อาจเลือกใช้สถิติ Dunnett's T3, Dunnett's C และ Games-Howell เป็นต้น

ค. เมื่อต้องการเปรียบเทียบค่าเฉลี่ยของกลุ่มควบคุม(Control Group) กับของกลุ่มทดลอง(Experimental Group) อาจเลือกใช้ สถิติ Dunnett หรือ Waller-Duncan เป็นต้น

ชนิดของการเปรียบเทียบ (Types of Comparison)

การเปรียบเทียบสามารถแบ่งออกได้หลายอย่าง (Atzenbeck, 2010; ESF, 2010; Maxwell and Delaney, 2010; Field, 2000) เช่น

1. **การเปรียบเทียบก่อนการทดสอบรวม (Prior Comparison)** ได้แก่การเปรียบเทียบค่าเฉลี่ยของกลุ่มพลวิจัยบางกลุ่มตามสมมติฐานหรือทฤษฎี ก่อนที่จะทำการทดสอบรวมว่าค่าเฉลี่ยของกลุ่มต่างๆทั้งหมดแตกต่างกันอย่างมีนัยสำคัญหรือไม่ หรือผู้วิจัยไม่สนใจที่จะรู้ประเด็นนี้ บางครั้งอาจเรียกว่า **การเปรียบเทียบตามแผนที่วางไว้ (Planned Comparison)** หรือ **การเปรียบเทียบความต่าง (Contrast Comparison)** หรือ **การวิเคราะห์ความต่าง (Contrast Analysis)**

2. **การเปรียบเทียบหลังการทดสอบรวม (Posteriori Comparison)** ได้แก่การเปรียบเทียบค่าเฉลี่ยของกลุ่มพลวิจัยบางกลุ่มหรือทั้งหมดตามสมมติฐาน ภายหลังจากที่ได้ทำการทดสอบรวมว่าค่าเฉลี่ยของกลุ่มต่างๆนั้นแตกต่างกันอย่างมีนัยสำคัญ และบางครั้งอาจเรียกว่า **การเปรียบเทียบภายหลัง (Post Hoc Comparison)** หรือ**การเปรียบเทียบที่ไม่ได้วางแผนไว้ (Unplanned Comparison)**

3. **การเปรียบเทียบรายคู่ (Pair Comparison)** ได้แก่การเปรียบเทียบค่าเฉลี่ยของพลวิจัยทีละคู่

4. **การเปรียบเทียบหลากหลาย (Multiple Comparison)** ได้แก่การเปรียบเทียบค่าเฉลี่ยของพลวิจัยทีละคู่สลับกันไปมาจนกระทั่งครบทุกคู่

5. **การเปรียบเทียบซับซ้อน (Complex Comparison)** ได้แก่การเปรียบเทียบค่าเฉลี่ยของพลวิจัยบางกลุ่มกับค่าเฉลี่ยรวม

ของบางกลุ่มในลักษณะที่ซับซ้อน เช่น $H_0 : 2\mu_1 = \frac{1}{2}\mu_2 + \frac{3}{4}\mu_3$ เป็นต้น

6. การเปรียบเทียบอิสระ (Orthogonal Comparison) ได้แก่การเปรียบเทียบค่าเฉลี่ยของพลวิจยทีละคู่อย่างเป็นอิสระ โดยไม่มีค่าเฉลี่ยใดได้รับการเปรียบเทียบมากกว่าหนึ่งครั้ง เช่น ในกรณีที่มี 4 กลุ่ม คือ A, B, C และ D การเปรียบเทียบดังกล่าวได้แก่การเปรียบเทียบ A-B และ C-D เท่านั้น

7. การเปรียบเทียบไม่เป็นอิสระ (Non-Orthogonal Comparison) ได้แก่การเปรียบเทียบค่าเฉลี่ยของพลวิจยทีละคู่โดยมีค่าเฉลี่ยของบางกลุ่มได้รับการเปรียบเทียบมากกว่าหนึ่งครั้ง เช่น ในกรณีที่มี 3 กลุ่ม คือ A, B และ C การเปรียบเทียบดังกล่าวได้แก่การเปรียบเทียบ A-B, A-C และ B-C เป็นต้น โดยแต่ละกลุ่มได้รับการเปรียบเทียบ 2 ครั้ง

8. การเปรียบเทียบค่าพิสัย (Range Comparison) ได้แก่การเปรียบเทียบค่าเฉลี่ยของพลวิจยหลายกลุ่มโดยอาศัยค่าพิสัยของความแตกต่างของค่าเฉลี่ย (คือค่าที่แตกต่างกันระหว่างค่าเฉลี่ยมากที่สุดกับค่าเฉลี่ยที่น้อยที่สุด) เป็นเกณฑ์ เช่น ในกรณีที่มี 4 กลุ่ม คือ A, B, C และ D เมื่อเรียงลำดับของค่าเฉลี่ยจากมากที่สุดไปหาน้อยที่สุดแล้ว และพบว่า A-D แตกต่างกับ B และ C อาจจะแตกต่างจาก A หรือ D หากว่ามีค่าเฉลี่ยน้อยกว่าค่าพิสัยระหว่าง A-D เป็นต้น การเปรียบเทียบแบบนี้บางครั้งเรียกว่า การเปรียบเทียบหลายขั้นตอน (Step-down Comparison)

ชนิดของความคลาดเคลื่อนในการเปรียบเทียบ (Types of Error in Comparison)

ปกติแล้วเนื่องจากการเปรียบเทียบดังกล่าวข้างต้นต้องทำหลายครั้ง จึงทำให้มี

โอกาสเกิดความคลาดเคลื่อนทางสถิติได้หลายอย่าง (Wilson, 2010; Wikipedia, 2010; Gold, 2010; Garson, 2009) เช่น

1. อัตราความคลาดเคลื่อนต่อการเปรียบเทียบแต่ละครั้ง (Per Comparison Error Rate: PC) ได้แก่ความผิดพลาดชนิดที่ 1 (Type I Error) อย่างหนึ่ง [ความผิดพลาดชนิดที่ 1 คือ โอกาสที่ผู้วิจัยจะปฏิเสธสมมติฐานศูนย์ (Null Hypothesis) ที่เป็นจริง] ที่เกิดขึ้นต่อการเปรียบเทียบค่าเฉลี่ยแต่ละครั้ง คือ

$$\alpha_{PC} = \frac{\alpha * n_{sig}}{n_{comp}}$$

เมื่อ

α_{PC} = อัตราความคลาดเคลื่อนต่อการเปรียบเทียบแต่ละครั้ง

α = ระดับความมีนัยสำคัญที่ผู้วิจัยกำหนด เช่น 0.05

n_{sig} = จำนวนครั้งของการเปรียบเทียบที่พบว่าค่าเฉลี่ยของกลุ่มพลวิจยแตกต่างกัน

n_{comp} = จำนวนครั้งของการเปรียบเทียบ

เช่น ถ้ามีกลุ่มพลวิจยที่ต้องการเปรียบเทียบ 3 กลุ่ม ผู้วิจัยทำการเปรียบเทียบ 3 ครั้ง พบว่า 2 ครั้ง (หรือ 2 คู่) ค่าเฉลี่ยมีความแตกต่างกันอย่างมีนัยสำคัญ ดังนั้นอัตราความคลาดเคลื่อนต่อการเปรียบเทียบแต่ละครั้งเมื่อผู้วิจัยกำหนดระดับความมีนัยสำคัญให้ $\alpha = 0.05$ คือ $\alpha_{PC} = \frac{0.05 * 2}{3} = 0.033$

เป็นต้น

2. อัตราความคลาดเคลื่อนต่อการเปรียบเทียบแต่ละชุด (Family-wise Error Rate: FW) ได้แก่ความผิดพลาดชนิดที่ 1 อีกอย่างหนึ่งเมื่อสมมติฐานศูนย์เป็นจริงแต่ได้รับการปฏิเสธในขณะที่ผู้วิจัยเปรียบเทียบค่าเฉลี่ยของตัวแปรต้น (Independent

Variable) ในแต่ละมิติสำหรับการวิเคราะห์ด้วย k-way ANOVA หรือ k-way ANCOVA กล่าวคือ

$$\alpha_{FW} = \frac{\alpha * n_{sig}}{n_{Fam}}$$

เมื่อ

α_{FW} = อัตราความคลาดเคลื่อนต่อการเปรียบเทียบแต่ละชุด

α = ระดับความมีนัยสำคัญที่ผู้วิจัยกำหนด เช่น 0.05

n_{sig} = จำนวนชุดที่มีการเปรียบเทียบแล้วพบว่าค่าเฉลี่ยของกลุ่มพลวิจัยแตกต่างกัน

n_{Fam} = จำนวนชุด หรือมิติการเปรียบเทียบ

เช่น งานวิจัยหนึ่งศึกษาสัมฤทธิ์ผลของวิธีสอน 3 วิธีของกลุ่มพลวิจัย 3 ระดับความสามารถ งานวิจัยนี้เรียกได้ว่ามีตัวแปรต้น 2 ชุด (Family) หากว่าผู้วิจัยต้องการเปรียบเทียบผลสัมฤทธิ์เฉพาะด้านวิธีสอนเท่านั้น อัตราความคลาดเคลื่อนดังกล่าวก็คือ

$$\alpha_{FW} = \frac{0.05 * 1}{2} = 0.025$$

3. อัตราความคลาดเคลื่อนต่อการทดลอง (Experiment-wise Error Rate : EW) ได้แก่ความคลาดเคลื่อนชนิดที่ 1 อีกอย่างหนึ่งก็เกิดขึ้นจากการทดลองเมื่อมีการเปรียบเทียบค่าเฉลี่ยของกลุ่มพลวิจัยแล้วพบว่ามีความแตกต่างกันอย่างน้อย 1 คู่ กล่าวคือ

$$\alpha_{EW} = \frac{\alpha * n_{sig}}{n_{Exp}}$$

เมื่อ

α_{EW} = อัตราความคลาดเคลื่อนต่อการทดลอง

α = ระดับความมีนัยสำคัญที่ผู้วิจัยกำหนด เช่น 0.05

n_{sig} = จำนวนการทดลองที่มีการเปรียบเทียบแล้วพบว่าค่าเฉลี่ยของกลุ่มพลวิจัยแตกต่างกัน

n_{Exp} = จำนวนการทดลองทั้งหมด

ตัวอย่าง เช่น มีวิธีสอน 3 วิธี เพื่อใช้สอน การอ่านเข้าใจความภาษาอังกฤษ ผู้วิจัยตั้งสมมติฐานแล้วมีการเปรียบเทียบผลการวิจัยของพลวิจัย 3 กลุ่ม พบว่าแตกต่างกันอย่างมีนัยสำคัญ ที่ระดับ $\alpha = 0.05$ ดังนั้นอัตราความคลาดเคลื่อนต่อการทดลอง คือ

$$\alpha_{EW} = \frac{0.05 * 1}{1} = 0.05$$

ดังนั้นจะเห็นได้ว่าในกรณีที่ทำการวิเคราะห์ข้อมูลโดยวิธี One-way ANOVA ค่า α_{EW} จะเท่ากับ α_{FW} เพราะจำนวนชุด (Family) = 1

การกำหนดระดับอัตราความคลาดเคลื่อน (Error Rate) ที่เหมาะสม

ในการกำหนดอัตราความคลาดเคลื่อนต่างๆที่เกิดจากการเปรียบเทียบ ผู้วิจัยควรพิจารณา ดังนี้ (SSTARS, 2009; Karpinski, 2006)

1. ในกรณีที่ทำการวิเคราะห์ข้อมูลด้วย One-Way ANOVA ผู้วิจัยควรที่จะเลือกสถิติเพื่อการเปรียบเทียบค่าเฉลี่ยที่สามารถควบคุมอัตราความคลาดเคลื่อนต่อการเปรียบเทียบแต่ละชุด (Family-wise Error Rate หรือ α_{FW}) ได้ แล้วจึงกำหนดระดับ $\alpha = 0.05$ สำหรับการทดลองรวม (Omnibus Test) และ $\alpha_{FW} = 0.05$

2. ในกรณีที่วิเคราะห์ข้อมูลด้วย k-way ANOVA ผู้วิจัยควรที่จะเลือกสถิติเพื่อการเปรียบเทียบค่าเฉลี่ยที่สามารถควบคุมอัตราความคลาดเคลื่อนต่อการทดลอง (Experiment-wise Error Rate หรือ α_{EW})

ได้เพราะหากว่าเลือกใช้สถิติที่ควบคุมได้ เฉพาะ α_{FW} เมื่อมีการเปรียบเทียบหลายมิติ [หรือหลายชุด (Family)] จะทำให้ค่า $\alpha_{EW} > 0.05$

ดังนั้น ในการเลือกใช้สถิติต่างๆเพื่อการเปรียบเทียบค่าเฉลี่ยของคะแนนรายคู่ หรือ หลายหลายคู่ พร้อมๆกันในการเปรียบเทียบก่อน (Priori Comparison) หรือหลัง (Posteriori Comparisons) การทดสอบรวม (Omnibus Test) ผู้วิจัยควรต้องคำนึงถึงเรื่องการควบคุม อัตราความคลาดเคลื่อนของ α_{FW} หรือ α_{EW} ด้วย กล่าวคือ ควรเลือกสถิติที่สามารถควบคุมให้ α_{FW} หรือ $\alpha_{EW} = \alpha = 0.05$ ในกรณีที่เป็น การวิจัยทาง สังคม ศาสตร์ หรือ มนุษยศาสตร์ สถิติใดที่ทำให้ $\alpha_{EW} < 0.05$ เรียกว่า สถิติ ทด สอบ เชิง อ นุ รักษ์ (Conservative Statistical Test) เช่น Scheffe ส่วนสถิติใดที่ทำให้ $\alpha_{EW} > 0.05$ เรียกว่าสถิติ ทดสอบเชิงก้าวหน้า (Liberal Statistical Test) เช่น Tukey Test เป็นต้น ดังนั้น หากเลือกได้ระหว่างสถิติ 2 ชนิดนี้ ผู้วิจัยควรเลือกใช้สถิติเชิงอนุรักษ์จะดีกว่าเพราะมี โอกาสทำให้เกิดความผิดพลาดชนิดที่ 1 (Type I Error) น้อยกว่าสถิติเชิงก้าวหน้า เว้นแต่ว่าผู้วิจัยจะต้องการทดสอบว่าความแตกต่างระหว่างค่าเฉลี่ยของผลวิจัยบางคู่มี โอกาสจะเกิดขึ้นได้หรือไม่

การเลือกใช้สถิติเพื่อการเปรียบเทียบอย่างเหมาะสม

เนื่องจากสถิติเพื่อใช้ในการเปรียบเทียบ มีเป็นจำนวนมาก และแต่ละอย่างมีจุดอ่อน และจุดแข็งแตกต่างกัน บางชนิดก็มีลักษณะเชิงอนุรักษ์ (Conservative) กล่าวคือ ทำให้ α_{EW} หรือ $\alpha_{FW} < 0.05$ [ทำให้เกิด Type I Error น้อยแต่ทำให้เกิด Type II Error มาก]

และบางชนิดก็มีลักษณะเชิงก้าวหน้า (Liberal) กล่าวคือ ทำให้ α_{EW} หรือ $\alpha_{FW} > 0.05$ [ทำให้เกิด Type I Error มาก แต่ทำให้เกิด Type II Error น้อย] สถิติเพื่อการเปรียบเทียบที่ดีควรจะสามารถควบคุมให้ α_{EW} หรือ $\alpha_{FW} = 0.05$ หรือใกล้เคียงกับระดับความมีนัยสำคัญที่ผู้วิจัยกำหนดไว้ให้มากที่สุด เกณฑ์ทั่วไปในการเลือกใช้สถิติเพื่อการเปรียบเทียบมีดังนี้ (SPSS 17.0, 2010; Prism Guide, 2010; Stevens, 1999)

1. เป็นการเปรียบเทียบก่อนหรือหลังการทดสอบรวม
2. ความแปรปรวน (σ^2) ของคะแนนของกลุ่มที่จะนำมาเปรียบเทียบกันเท่ากันหรือไม่
3. ขนาดของกลุ่มตัวอย่าง (n) ที่จะนำมาเปรียบเทียบกันเท่ากันหรือไม่
4. สถิติที่ใช้สามารถควบคุมให้ α_{EW} หรือ $\alpha_{FW} = 0.05$ หรือใกล้เคียงกับระดับความมีนัยสำคัญที่ผู้วิจัยกำหนดไว้ให้มากที่สุดหรือไม่
5. สถิติที่จะใช้คำนวณง่ายหรือไม่ หรือสามารถหาได้จากโปรแกรมสำเร็จรูปทางสถิติหรือไม่
6. สถิติที่จะใช้เหมาะสมกับลักษณะของข้อมูล (เช่น มาตรฐาน การกระจาย และขนาด) และวัตถุประสงค์ของการวิจัย (เช่น เพื่อเปรียบเทียบแบบรายคู่ แบบจับช้อน หรือแบบกลุ่มควบคุมกับกลุ่มทดลองเท่านั้น) หรือไม่

ดังนั้น ต่อไปนี้ผู้เขียนจะกล่าวถึงการเลือกใช้สถิติเพื่อการเปรียบเทียบผลการทดลองในการวิจัยให้เหมาะสมกับวัตถุประสงค์ของการวิจัยและลักษณะของข้อมูลการวิจัย

การเปรียบเทียบก่อนการทดสอบรวม (Prior Comparison)

การเปรียบเทียบชนิดนี้เลือกใช้ดังนี้ (Karpinski, 2006)

1. เมื่อผู้วิจัยต้องการวิเคราะห์ข้อมูลเพื่อการยืนยัน (Confirmatory Data Analysis) แนวคิด ทฤษฎี หรือความเชื่อที่มีมาก่อนโดยผู้วิจัยต้องการที่จะเปรียบเทียบค่าเฉลี่ยของคะแนนของพลวิจัยเพียงบางกลุ่มเท่านั้น เช่น กรณีที่มีกลุ่มควบคุม 1 กลุ่มคือกลุ่ม A และกลุ่มทดลอง 2 กลุ่ม คือ กลุ่ม B และ กลุ่ม C ผู้วิจัยประสงค์ที่จะเทียบคะแนนค่าเฉลี่ยของกลุ่มควบคุมกับคะแนนค่าเฉลี่ยของกลุ่ม B เท่านั้น โดยมีสมมติฐานการวิจัยจากวรรณกรรมที่เกี่ยวข้องว่า

$$H_0 : 2\mu_1 = 3/4\mu_2$$

2. ผู้วิจัยไม่ต้องการที่จะรู้ว่าคะแนนค่าเฉลี่ยของทั้ง 3 กลุ่มแตกต่างกันอย่างมีนัยสำคัญหรือไม่

ดังนั้น การเปรียบเทียบก่อนการทดสอบมีขั้นตอนดังนี้ คือ

1. ไม่ต้องทำการทดสอบรวม กล่าวคือไม่ต้องทำการทดสอบความแตกต่างของค่าเฉลี่ยของกลุ่มควบคุมและกลุ่มทดลองด้วย F-test
2. ดำเนินการเปรียบเทียบค่าเฉลี่ยเฉพาะของกลุ่มที่ต้องการเปรียบเทียบเท่านั้น โดยกำหนดค่าสัมประสิทธิ์การเปรียบเทียบของแต่ละกลุ่มซึ่งรวมกันแล้วจะต้องเท่ากับ 0 พอดี เช่น 1 กับ -1 ในกรณีที่เปรียบเทียบ 2 กลุ่ม หรือ 1 กับ -.50 และ -.50 ในกรณี 3 กลุ่ม หรือ .50, .50, -.50 และ -.50 ในกรณี 4 กลุ่ม เป็นต้น
3. เลือกสถิติที่สอดคล้องกับลักษณะของข้อมูล และวัตถุประสงค์ของการเปรียบเทียบ
4. เลือกสถิติที่มีอัตราความเคลื่อน (Error Rate) ที่ผู้วิจัยยอมเสี่ยงให้เกิดขึ้นได้ซึ่งมีอยู่

หลายชนิดดังกล่าวแล้วข้างต้น โดยเลือกใช้สถิติที่สามารถควบคุมให้ α_{EW} หรือ $\alpha_{FW} = 0.05$ หรือใกล้เคียงกับ ระดับความมีนัยสำคัญที่ผู้วิจัยกำหนดไว้ให้มากที่สุด

5. ในทางทฤษฎีแล้ว เมื่อนักวิจัยทำการเปรียบเทียบก่อนการทดสอบรวมแล้วก็ไม่ต้องทำการทดสอบภายหลังอีก และในทำนองเดียวกัน เมื่อทำการเปรียบเทียบภายหลังการทดสอบรวมแล้วก็ไม่มีความจำเป็นที่จะต้องทำการทดสอบก่อนอีก (Field, 2000)

สถิติที่นิยมใช้เพื่อการเปรียบเทียบก่อนการทดสอบรวม

สถิติที่นิยมใช้เพื่อการเปรียบเทียบก่อนการทดสอบรวมมีหลายอย่าง แต่ที่นิยมใช้กันอย่างแพร่หลายในโปรแกรมสำเร็จรูปทางสถิติ (Ready-made Statistical Package) ที่ใช้ทั่วไป เช่น SPSS (Statistical Package for the Social Sciences) SAS (Statistical Applications for Sciences) และ BMDP (Biological Medical Development Program) แต่ในบทความนี้ ผู้เขียนจะเน้นเฉพาะสถิติที่มีใช้ในโปรแกรม SPSS เท่านั้น เนื่องเป็นโปรแกรมที่นักวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์ทั่วไปนิยมใช้ซึ่งได้แก่สถิติต่อไปนี้ (Wikipedia, 2010; Garson, 2009; SSTARs, 2009)

1. t-test

สถิตินี้ควรเลือกใช้เมื่อมีการเปรียบเทียบค่าเฉลี่ยเพียงคู่เดียว เพราะสามารถควบคุมให้ α_{EW} หรือ $\alpha_{FW} = 0.05$ ได้ แต่หากมีการเปรียบเทียบมากคู่แล้วจะไม่สามารถควบคุมอัตราความคลาดเคลื่อนดังกล่าวได้

2. Bonferroni Test

สถิตินี้บางครั้งเรียกว่า Dunn Test และได้รับความนิยมมากเพราะสามารถใช้เพื่อการเปรียบเทียบก่อน และหลังการทดสอบรวมได้ สำหรับการเปรียบเทียบก่อน สถิตินี้ใช้เปรียบเทียบรายคู่แบบหลากหลาย (Multiple Comparison) ที่มีลักษณะเชิงอนุรักษ์มาก เพราะไม่สามารถควบคุมให้ α_{EW} หรือ $\alpha_{FW} = 0.05$ ได้ เนื่องจากอัตราความคลาดเคลื่อน = $0.05/p$ เมื่อ $p =$ จำนวนครั้งการเปรียบเทียบ ดังนั้นควรใช้เมื่อมีการเปรียบเทียบรายคู่จำนวนไม่มาก (เช่น 2-3 คู่)

3. Sidak Test

สถิตินี้พัฒนามาจาก Bonferroni Test บางครั้งเรียกกันว่า Dunn-Sidak Test (FSU, 2010) และมีลักษณะเชิงอนุรักษ์น้อยกว่า Bonferroni Test เพราะทำให้ค่า α_{EW} หรือ α_{FW} ใกล้เคียงกับค่า α ที่ผู้วิจัยกำหนดได้มากกว่า จึงเป็นสถิติที่เหมาะสมสำหรับการเปรียบเทียบรายคู่ของค่าเฉลี่ยของพลวิจัยมากกว่า Bonferroni Test

4. Dunnett Test

สถิตินี้พัฒนาโดย Dunnett ในปี 1955 และเหมาะสำหรับการเปรียบเทียบรายคู่ของค่าเฉลี่ยของพลวิจัยกลุ่มควบคุมกับกลุ่มทดลองเท่านั้น เพราะสามารถควบคุมให้ α_{EW} หรือ $\alpha_{FW} = 0.05$ ได้เป็นอย่างดี

การเปรียบเทียบหลังการทดสอบรวม (Posteriori Comparisons)

การเปรียบเทียบชนิดนี้เลือกใช้ดังนี้ (Karpinski, 2006)

1. เมื่อผู้วิจัยต้องการวิเคราะห์ข้อมูลเพื่อการสำรวจ (Exploratory Data Analysis)

กล่าวคือผู้วิจัยยังไม่มีแนวคิด ทฤษฎี หรือความเชื่อมาก่อนว่าโดยเฉลี่ยแล้วคะแนนของพลวิจัยบางคู่แตกต่างกัน แต่ก็มีสมมติฐานบางอย่างที่ต้องจะทดสอบว่าเป็นจริงหรือไม่ เช่น กรณีที่มีกลุ่มควบคุม 1 กลุ่มเรียนภาษาอังกฤษเพื่อการอ่านเข้าใจความภาษาอังกฤษ โดยวิธีประเพณีนิยม (Traditional Instruction) และมีกลุ่มทดลอง 2 กลุ่มเรียนโดยวิธีใช้เครื่องคอมพิวเตอร์ช่วยสอน (Computer-Assisted Instruction) และวิธีใช้ระบบเครือข่ายคอมพิวเตอร์ (Web-based Instruction) ที่ผู้วิจัยต้องการที่จะทราบว่าการเรียนดังกล่าววิธีใดจะได้ผลดีกว่ากัน จึงตั้งสมมติฐานขึ้นเพื่อทำการเปรียบเทียบรายคู่ของผลการเรียนแต่ละวิธี เป็นต้น

2. ผู้วิจัยต้องการที่จะรู้ว่าโดยเฉลี่ยแล้วคะแนนของพลวิจัยในการทดลองแต่ละกลุ่มแตกต่างกันอย่างมีนัยสำคัญหรือไม่

ดังนั้นการเปรียบเทียบหลังการทดลองจึงมีขั้นตอนดังนี้ (Garson, 2009)

1. ทำการทดสอบรวม (Omnibus Test) ว่าโดยเฉลี่ยแล้ว กลุ่มพลวิจัยมีคะแนนแตกต่างกันอย่างมีนัยสำคัญหรือไม่ โดยใช้ ANOVA หรือ ANCOVA แล้วแต่รูปแบบการวิจัย (Research Design) และลักษณะของข้อมูลเพื่อการวิจัย หากการทดสอบพบว่าคะแนนโดยเฉลี่ยของพลวิจัยแตกต่างกันอย่างไม่มีนัยสำคัญที่ระดับ α ที่กำหนดไว้ก็ไม่จำเป็นต้องวิเคราะห์อีกต่อไป

2. หากพบว่าคะแนนต่างกันอย่างมีนัยสำคัญก็ทำการเปรียบเทียบรายคู่ต่อไปโดยเลือกใช้สถิติที่สอดคล้องกับวัตถุประสงค์การวิจัย ลักษณะของข้อมูล และลักษณะของอัตราความคลาดเคลื่อนที่ผู้วิจัยยอมเสี่ยงให้เกิดขึ้นได้ดังได้กล่าวมาแล้วข้างต้น

สถิติที่นิยมใช้เพื่อการเปรียบเทียบหลังการทดสอบรวม

สถิติที่นิยมใช้เพื่อการเปรียบเทียบหลังการทดสอบรวมมีหลายอย่าง แต่ที่นิยมใช้กันอย่างแพร่หลายในโปรแกรมสำเร็จรูปทางสถิติอาจแบ่งได้ 2 กลุ่มใหญ่ เพื่อใช้ในกรณีที่แตกต่างกัน คือ

ก. เมื่อความแปรปรวนของคะแนนของกลุ่มพลวิจัยไม่แตกต่างกัน ($\delta_1^2 = \delta_2^2$)

ข. เมื่อความแปรปรวนของคะแนนของกลุ่มพลวิจัยแตกต่างกัน ($\delta_1^2 \neq \delta_2^2$)

ต่อไปนี้เป็นสถิติที่นิยมใช้ในแต่ละกรณีดังกล่าวแล้ว (Wikipedia, 2010; Garson, 2009; SSTARS, 2009; Karpinski, 2006)

ก. เมื่อความแปรปรวนของคะแนนของกลุ่มพลวิจัยไม่แตกต่างกัน ($\delta_1^2 = \delta_2^2$)

1. Least Significant Difference Test (LSD Test)

เป็นสถิติที่คิดค้นสูตรโดย Fisher ในปี 1993 และมีลักษณะเชิงก้าวหน้ามากที่สุด (Most liberal test) ในบรรดาสถิติที่ใช้เพื่อการเปรียบเทียบค่าเฉลี่ยรายคู่ จึงเหมาะสำหรับใช้ในกรณีต่อไปนี้

1. กลุ่มที่นำมาเปรียบเทียบกันควรมีจำนวนไม่เกิน 3 กลุ่ม เพราะจะทำให้ $\alpha_{EW} \leq 0.05$
2. ต้องการเปรียบเทียบค่าเฉลี่ยของทั้ง 3 กลุ่มในครั้งเดียวกัน
3. หากว่ามีกลุ่มตัวอย่างมากกว่า 3 กลุ่มไม่ควรใช้สถิตินี้ เพราะจะทำให้ $\alpha_{EW} > 0.05$

2. Honestly Significant Difference Test (HSD Test, Tukey–Kramer Test หรือ Tukey’s A Test)

เป็นสถิติที่คิดค้นสูตรโดย John Tukey เมื่อปี 1953 และเหมาะสำหรับใช้ในกรณีต่อไปนี้

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) หรือไม่เท่ากัน ($n_1 \neq n_2$) ก็ได้
2. ใช้เปรียบเทียบรายคู่จำนวนไม่มาก (จำนวน 2-3 คู่) และสามารถควบคุมให้ $\alpha_{EW} < 0.05$ ได้เป็นอย่างดี จึงถือได้ว่าเป็นสถิติเชิงอนุรักษ์
3. ในกรณีที่ขนาดกลุ่มตัวอย่างไม่เท่ากันจะเป็นสถิติเชิงอนุรักษ์มาก
4. หากว่ามีจำนวนคู่ในการเปรียบเทียบมากจะเป็นสถิติเชิงก้าวหน้า (Liberal) กล่าวคือ จะทำให้ $\alpha_{EW} > 0.05$

3. Student-Newman-Keults Test (SNK Test)

สถิตินี้คิดค้นโดยสูตรโดย Student ในปี 1927 Newman ในปี 1939 และ Keults ในปี 1952 แล้วได้มีการพัฒนาสูตรให้ดียิ่งขึ้นในปี 1952 และเหมาะสำหรับใช้ในกรณีดังต่อไปนี้

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$)
2. ใช้เปรียบเทียบรายคู่ และเปรียบเทียบหลายระดับ (Step-down comparison) โดยอาศัยค่าพิสัยของค่าเฉลี่ย
3. เป็นสถิติที่มีอำนาจในการเปรียบเทียบได้ดีกว่า Tukey’s A Test
4. เนื่องจากสถิตินี้ไม่สามารถควบคุมค่า α_{EW} ได้ และไม่ให้ค่าช่วงความเชื่อมั่น (Confidence Interval) ดังนั้น SNK จึงไม่เหมาะที่จะนำมาใช้หากสามารถเลือกใช้สถิติอื่นได้

4. REGW-Q Test

สถิตินี้คิดค้นโดย Ryan ในปี 1959 และปรับปรุงแก้ไขโดย Einot และ Gabriel ในปี 1975 และ Welsch ได้ปรับปรุงแก้ไขให้

ที่ความเหมาะสมยิ่งขึ้นในปี 1977 แต่เนื่องจากสถิตินี้อาศัยการแจกแจงแบบ Q (Q Distribution) จึงเรียกว่า REW-Q และเหมาะสำหรับใช้ในกรณีต่อไปนี้

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$)
2. การเปรียบเทียบรายคู่สามารถควบคุมให้ $\alpha_{EW} \leq 0.05$ และมีอำนาจในการเปรียบเทียบสูงกว่า HSD ของ Tukey

5. Tukey's B Test (Wholly Significant Difference Test หรือ WSD Test)

สถิตินี้พัฒนาขึ้นโดย John Tukey เพื่อแก้ปัญหาของสถิติ Turkey's A ที่ทำให้ $\alpha_{EW} < 0.05$ และสถิติ SNK ที่ไม่สามารถจะควบคุมค่า α_{EW} ที่เกิดจากการเปรียบเทียบได้ โดยการนำค่าวิกฤติ (Critical Value) ที่ได้จากสถิติทั้ง 2 ชนิดมาหาค่าเฉลี่ยกันและใช้เป็นค่าวิกฤติของ Tukey's B Test

แต่เนื่องจากการคำนวณหาค่าวิกฤติของสถิตินี้ยุ่งยากมากและค่า α_{EW} ที่ได้ไม่แน่นอน เนื่องจากไม่สามารถควบคุมค่าดังกล่าวจากสถิติ SNK ได้ ดังนั้น หากสามารถใช้สถิติอื่นได้ นักวิจัยจึงไม่ควรใช้สถิตินี้เพื่อ การเปรียบเทียบค่าเฉลี่ยรายคู่ของผลการวิจัยภายหลังจากการทดสอบรวม

6. Scheffe Test

สถิตินี้พัฒนาขึ้นโดย Henry Scheffe ในปี 1953 โดยการพัฒนาศูตรการวิเคราะห์เปรียบเทียบจาก Tukey's A Test สถิตินี้มีลักษณะเชิงอนุรักษ์มากที่สุด (Most conservative test) ในบรรดาสถิติที่ใช้เพื่อการเปรียบเทียบ (Pierce, 2003) และเหมาะสมกับการใช้ในกรณีต่อไปนี้

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) หรือไม่เท่ากัน ($n_1 \neq n_2$) ก็ได้
2. เมื่อต้องการเปรียบเทียบค่าเฉลี่ยรายคู่ (Pair Comparison) และการเปรียบเทียบแบบซับซ้อน (Complex Comparison)
3. เมื่อต้องการควบคุมขนาดของ $\alpha_{EW} \leq 0.05$ ของทุกการเปรียบเทียบ
4. ควรจะมีกลุ่มพลวิจัยมากกว่า 3 กลุ่ม เพราะว่ามีจำนวนกลุ่มพลวิจัยมีน้อยกว่า 4 กลุ่มอำนาจในการวิเคราะห์เปรียบเทียบของสถิตินี้มีน้อยกว่าสถิติอื่น (Neel, 1997)

อนึ่ง เป็นที่น่าสังเกตว่า เมื่อใช้ข้อมูลชุดเดียวกัน Scheffe Test จะมีความไวหรืออำนาจในการเปรียบเทียบน้อยกว่า Tukey's HSD Test ซึ่งมีลักษณะก้าวหน้า (Liberal) ด้วยเหตุนี้ Scheffe Test จึงได้รับความนิยมมากสำหรับนักวิจัยที่ต้องการให้เกิด Type I Error ในการวิจัยน้อยที่สุด (Pierce, 2003)

7. Bonferroni Test

Bonferroni Test นอกจากจะใช้เพื่อการเปรียบเทียบก่อนการทดสอบรวมแล้ว ยังสามารถใช้เพื่อการเปรียบเทียบภายหลังการทดสอบรวมด้วย สถิตินี้ Carlo Emilio Bonferroni นักคณิตศาสตร์ชาวอิตาลี เป็นคนคิดค้นขึ้นเป็นคนแรก โดยอาศัยแนวคิดพื้นฐานมาจาก student's t test และบางครั้งเรียกว่า Dunn's Test เพราะ Dunn เป็นผู้พัฒนาสูตรขึ้นใช้ในปี 1961 สถิตินี้ใช้กันอย่างกว้างขวางที่สุดในบรรดาสถิติที่ใช้เพื่อการเปรียบเทียบค่าเฉลี่ยรายคู่ภายหลังการทดสอบรวมแล้ว (Newman, 2009) ทั้งนี้เพราะมีความยืดหยุ่นสูง กล่าวคือ การคำนวณทำได้ง่าย และใช้ได้กับสถิติอื่นด้วยนอกจากใช้ค่าเฉลี่ยแล้ว เช่น ใช้ค่าสัมประสิทธิ์สหสัมพันธ์ ค่า t-test และ

ค่า F-test ก็ได้ แต่ว่าสถิตินี้มีแนวโน้มว่าจะมีอำนาจ (power) ในการทดสอบต่ำ เพราะ

1. อาศัยแนวคิดที่ว่าสมมติฐานศูนย์เป็นจริง ซึ่งเป็นเรื่องเป็นไปได้ยากเมื่อมีการทดสอบรวม (Omnibus Test) และมีการปฏิเสธ (แสดงว่าอย่างน้อยต้องมีค่าเฉลี่ยของกลุ่มพลวิจัย 2 กลุ่มที่แตกต่างกันอย่างมีนัยสำคัญ) ซึ่งในกรณีนี้จะทำให้ $\alpha_{FW} > 0.05$

2. เมื่อการเปรียบเทียบรายคู่ที่ไม่เป็นอิสระ (Non-Orthogonal Comparison) เช่น กรณีที่มี 3 กลุ่มคือ A, B และ C แล้วเปรียบเทียบ A-B, A-C และ B-C

3. การทดสอบแต่ละครั้งมีความผิดพลาดแบบ Type II สูงเนื่องจาก Bonferroni Test พยายามแก้ไขความผิดพลาดแบบ Type I มากเกินไป

ดังนั้น นักวิจัยควรเลือกใช้ Bonferroni Test สำหรับกรณีดังนี้

1. เมื่อมีจำนวนกลุ่มพลวิจัยมาก Bonferroni Test อาจไม่มีอำนาจมากเพียงพอที่จะทำให้การเปรียบเทียบพบความแตกต่างอย่างมีนัยสำคัญได้ ดังนั้น หากต้องการควบคุมขนาดของ α_{FW} ให้อยู่ในระดับ 0.05 ของทุกๆ การเปรียบเทียบก็ควรเลือกใช้สถิตินี้เมื่อมีกลุ่มพลวิจัยไม่เกิน 5 กลุ่ม และเป็นการเปรียบเทียบอิสระ (Prism Guide, 2010)

2. Bonferroni Test เป็นสถิติที่อนุรักษ์มากเมื่อใช้กับกลุ่มพลวิจัยจำนวนมาก กล่าวคือระดับ $\alpha_{FW} < 0.05$ ดังนั้น ควรใช้เมื่อมีกลุ่มตัวอย่างไม่เกิน 5 กลุ่ม

8. Sidak Test (Sidak-Bonferroni Test หรือ Dunn-Sidak Test)

สถิตินี้เป็นสูตรที่ปรับปรุงมาจาก Bonferroni Test เพื่อแก้ปัญหาเรื่องค่า $\alpha_{FW} < 0.05$ ที่มี

ลักษณะเชิงอนุรักษ์มาก โดยแก้ไขสูตรการคำนวณเล็กน้อย (Newsom, 2009)

สถิตินี้ควรเลือกใช้ดังนี้

1. เมื่อต้องการควบคุมขนาดของ α_{FW} ให้คงที่หรือน้อยกว่า α ที่กำหนดไว้ (เช่น $\alpha_{FW} < 0.05$)

2. เมื่อต้องการสถิติที่เชิงก้าวหน้า (Liberal Test) กว่า Bonferroni เล็กน้อย

9. Hochberg's GF2 Test

สถิตินี้ปรับปรุงมาจากสถิติ Tukey's A Test และเหมาะสมสำหรับการเปรียบเทียบค่าเฉลี่ยของกลุ่มพลวิจัยดังต่อไปนี้

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) หรือไม่เท่ากัน ($n_1 \neq n_2$) ก็ได้

2. ต้องการควบคุมขนาดของ α_{FW} ไม่ให้มากกว่า α ที่กำหนด

3. มีลักษณะเชิงอนุรักษ์มากกว่า Tukey's A เมื่อกลุ่มพลวิจัยมีขนาดเท่ากัน

4. มีลักษณะเชิงอนุรักษ์มากกว่า Tukey's B หรือ Tukey-Kramer Test เมื่อกลุ่มพลวิจัยมีขนาดต่างกัน

10. Gabriel Test

สถิตินี้ใช้เพื่อการเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างที่มีลักษณะดังต่อไปนี้

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) หรือไม่เท่ากัน ($n_1 \neq n_2$) ก็ได้

2. ในกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากัน ($n_1 = n_2$) ค่าของ Gabriel Test จะเท่ากับค่าของ Hochberg's GF2 Test พอดีเมื่อวิเคราะห์ด้วย ANOVA

3. ในกรณีที่กลุ่มตัวอย่างมีขนาดแตกต่างกัน ($n_1 \neq n_2$) ค่าของสถิตินี้จะมีลักษณะเชิงอนุรักษ์น้อยกว่า Hochberg's GF2 Test

11. Duncan Test

Duncan Test เป็นสถิติที่ใช้ทดสอบการเปรียบเทียบหลายขั้นตอน (Step-down Test) หรือ สถิติเชิงพิสัย (Range Test) ซึ่งมีลักษณะเฉพาะดังนี้ คือ

1. สถิติจะทำการเปรียบเทียบตามค่าเฉลี่ยของกลุ่มพลวิจัยต่างๆว่าแตกต่างกันอย่างมีนัยสำคัญหรือไม่ โดยอาศัยสมมติฐานศูนย์ (Null Hypothesis)

2. หากทดสอบดังกล่าวแล้วพบว่าค่าเฉลี่ยของกลุ่มต่างๆแตกต่างกันอย่างไม่มีนัยสำคัญ การทดสอบจะสิ้นสุดลง

3. แต่หากว่าการทดสอบในขั้นที่ 1 พบว่าค่าเฉลี่ยของบางกลุ่มแตกต่างกันอย่างมีนัยสำคัญ สถิติจะทำการทดสอบต่อไป

4. สถิติทำการวิเคราะห์เปรียบเทียบว่าค่าเฉลี่ยของพลวิจัยกลุ่มใดแตกต่างกัน และกลุ่มใดไม่แตกต่างกันอย่างมีนัยสำคัญ

5. สถิตินี้ไม่มีค่าช่วงความเชื่อมั่น (Confidence Interval)

6. บางครั้งการจัดกลุ่มพลวิจัยที่มีค่าเฉลี่ยไม่แตกต่างกันอาจจะเข้าซ้อนกันได้ กล่าวคือพลวิจัยกลุ่มหนึ่งอาจจะถูกจัดกลุ่มให้มีความเฉลี่ยอยู่ใน 2 กลุ่มที่มีความเฉลี่ยแตกต่างกันอย่างไม่มีนัยสำคัญ

ดังนั้น เพื่อหลีกเลี่ยงปัญหาดังกล่าว โดยเฉพาะในข้อ 5 และ 6 ข้างต้น หากเลือกใช้สถิติอื่นเพื่อการเปรียบเทียบค่าเฉลี่ยเป็นรายคู่ได้ นักวิจัยไม่ควรเลือกใช้สถิติที่มีลักษณะการทดสอบเปรียบเทียบหลายขั้นตอน เช่น REGW-Q, REGW-F และ NSK เป็นต้น โดยเฉพาะอย่างยิ่งเมื่อขนาดกลุ่มตัวอย่างแตกต่างกัน

Duncan Test เหมาะสำหรับข้อมูลที่มีลักษณะดังต่อไปนี้คือ

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) หรือไม่เท่ากัน ($n_1 \neq n_2$) ก็ได้
2. เมื่อไม่ต้องการควบคุมขนาดของ α_{FW} เพราะสถิตินี้ทำไม่ได้

12. REGW-F Test

สถิตินี้คิดค้นขึ้นมาโดยนักสถิติ 4 ท่านคือ Ryan, Einot, Gabriel และ Welsch มีจำนวน 2 สูตร คือ REGW-F จากพื้นฐานของ F-Test และ REGW-F จากพื้นฐานของ Studentized Range Test สำหรับ REGW-F Test เหมาะสำหรับข้อมูลและการทดสอบที่มีลักษณะดังต่อไปนี้

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$)
2. เมื่อต้องการสถิติเปรียบเทียบหลายขั้นตอน (Step-down Test)
3. เป็นสถิติที่มีอำนาจในการเปรียบเทียบมากกว่า Duncan Test และ SNK Test

13. Dunnett Test

Dunnett Test เหมาะสำหรับการทดสอบเพื่อการเปรียบเทียบภายหลังของกลุ่มตัวอย่างที่ลักษณะดังต่อไปนี้

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) หรือไม่เท่ากัน ($n_1 \neq n_2$) ก็ได้
2. ใช้เปรียบเทียบค่าเฉลี่ยของกลุ่มควบคุมกับกลุ่มทดลองเท่านั้น
3. สามารถควบคุมขนาด $\alpha_{FW} = \alpha$ ได้พอดีของการเปรียบเทียบ

14. Waller-Duncan Test

สถิตินี้พัฒนามาจาก Bayesian Approach และเป็นสถิติเชิงพิสัย (Range Test) หรือ สถิติเปรียบเทียบหลายขั้นตอน (Step-down Test) ดังได้กล่าวมาแล้ว และเหมาะสำหรับ

การทดสอบการเปรียบเทียบดังต่อไปนี้ (SPSS 17.0, 2010)

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) หรือไม่เท่ากัน ($n_1 \neq n_2$) ก็ได้
2. ผู้วิจัยต้องการกำหนดอัตราส่วนความสัมพัทธ์ของ ความเสียหาย (Relative Seriousness) ระหว่างความคลาดเคลื่อนของ Type I Error กับ Type II Error เอง แทนที่จะกำหนดเป็นระดับความมีนัยสำคัญ กล่าวคือ ถ้ากำหนดให้ $k = 50, 100$ และ 500 เมื่อ $k =$ อัตราส่วน ผลที่ได้จะเทียบเท่าระดับ $\alpha = 0.10, 0.05$ และ 0.01 โดยประมาณ

ข. เมื่อความแปรปรวนของคะแนนของกลุ่มพลวิชัยแตกต่างกัน ($\delta_1^2 \neq \delta_2^2$)

ในกรณีที่ความแปรปรวน (δ^2) ของคะแนนของกลุ่มพลวิชัยที่นำมาเปรียบเทียบกันแตกต่างกันอย่างมีนัยสำคัญ ($\delta_1^2 \neq \delta_2^2$) ซึ่งอาจจะเกิดจากขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) หรือไม่เท่ากัน ($n_1 \neq n_2$) ก็ได้ นักวิจัยสามารถเลือกใช้สถิติเพื่อการเปรียบเทียบได้หลายอย่าง เช่น

1. Tamhane's T2 Test
2. Games-Howell Test
3. Dunnett's T3 Test
4. Dunnett's C Test

สถิติแต่ละอย่างเหมาะสำหรับใช้กับข้อมูล และวัตถุประสงค์ที่มีลักษณะเฉพาะดังต่อไปนี้ (Garson, 2009)

1. Tamhane's T2 Test

สถิตินี้เหมาะสำหรับใช้กับข้อมูล และวัตถุประสงค์ดังต่อไปนี้

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) หรือไม่เท่ากัน ($n_1 \neq n_2$) ก็ได้

2. ใช้เพื่อการเปรียบเทียบรายคู่ (Pair Comparison) โดยอาศัยแนวคิดจาก t-test

3. ถ้าขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) จะเป็นสถิติเชิงอนุรักษ์ (Conservative Test) กล่าวคือ ทำให้ $\alpha_{FW} < \alpha$ ที่กำหนด

4. ถ้า $n_1 \neq n_2$ สถิตินี้จะมีอำนาจการเปรียบเทียบสูงกว่า Tukey's A Test

5. ในกรณีกลุ่มตัวอย่างมีขนาดต่างกันมาก ค่าของ α_{FW} อาจจะมีมากกว่า α ที่กำหนดได้ ซึ่งจะทำให้สถิตินี้มีลักษณะก้าวหน้า (Liberal Test)

2. Games-Howell Test

สถิตินี้เหมาะสำหรับข้อมูลที่มีลักษณะดังนี้

1. ความแปรปรวนของคะแนนของกลุ่มพลวิชัยเท่ากันหรือแตกต่างกัน ($\delta_1^2 = \delta_2^2$ หรือ $\delta_1^2 \neq \delta_2^2$) ก็ได้

2. ขนาดของกลุ่มพลวิชัยเท่ากันหรือแตกต่างกัน ($n_1 = n_2$ หรือ $n_1 \neq n_2$) ก็ได้

3. เมื่อกลุ่มตัวอย่างมีขนาดเล็ก สถิตินี้จะมีลักษณะก้าวหน้า (Liberal) กล่าวคือจะทำให้ $\alpha_{FW} > \alpha$ ที่กำหนด ดังนั้นควรเลือกใช้เมื่อขนาดตัวอย่างในแต่ละกลุ่ม > 5

4. เป็นสถิติที่มีอำนาจในการเปรียบเทียบสูงกว่า Dunnett's T3 Test และ Dunnett's C Test

5. เมื่อผู้วิจัยไม่ทราบความแปรปรวนของกลุ่มพลวิชัยที่นำมาเปรียบเทียบกันว่าเป็นอย่างไร

3. Dunnett's T3 Test

สถิตินี้พัฒนามาจากพื้นฐานของการเปรียบเทียบรายคู่ของ Studentized Maximal Module และเหมาะสำหรับข้อมูลที่มีลักษณะดังนี้

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) หรือไม่เท่ากัน ($n_1 \neq n_2$) ก็ได้
2. เมื่อกลุ่มตัวอย่างแตกต่างกันมาก ($n_1 \neq n_2$) มีอำนาจในการเปรียบเทียบน้อยกว่า Games-Howell Test
3. Dunnett's T3 Test มีอำนาจในการเปรียบเทียบน้อยกว่า Dunnett's C Test เมื่อ df ของ MS_{error} มีขนาดใหญ่
4. สามารถควบคุมให้ $\alpha_{FW} = \alpha$ ที่กำหนดได้

4. Dunnett's C Test

สถิตินี้พัฒนามาจากพื้นฐานของการเปรียบเทียบรายคู่ของ Studentized Range Test ซึ่งมีการเปรียบเทียบหลายขั้นตอน (Step-down Test) และเหมาะสมสำหรับข้อมูลดังต่อไปนี้

1. เมื่อความแปรปรวนของคะแนนของกลุ่มพลวิชัยแตกต่างกัน ($\sigma_1^2 \neq \sigma_2^2$) และขนาดของกลุ่มตัวอย่างแตกต่างกัน ($n_1 \neq n_2$)
2. เมื่อกลุ่มตัวอย่างแตกต่างกันมาก ($n_1 \neq n_2$) มีอำนาจในการเปรียบเทียบน้อยกว่า Games-Howell Test
3. Dunnett's C Test มีอำนาจในการเปรียบเทียบมากกว่า Dunnett's T3 Test เมื่อ df ของ MS_{error} มีขนาดใหญ่
4. สามารถควบคุมให้ $\alpha_{FW} = \alpha$ ที่กำหนดได้

ข้อเสนอแนะทั่วไปในการใช้สถิติเพื่อการเปรียบเทียบ

เนื่องจากสถิติเพื่อการเปรียบเทียบความแตกต่างของค่าเฉลี่ยมีมาก และแต่ละอย่างมีข้อดีและข้อเสียต่างกัน รวมทั้งใช้เพื่อวัตถุประสงค์ที่ต่างกันดังได้กล่าวมาแล้ว ดังนั้น เพื่อความสะดวกในการใช้งาน ได้มีนักวิจัยทำการศึกษาเปรียบเทียบการใช้สถิติ

ต่างๆ เหล่านี้และมีข้อเสนอแนะทั่วไปในการใช้ดังนี้ (Garson, 2009; SSTARs, 2009; Field, 2000; Becker, 1999)

1. เมื่อวิเคราะห์ข้อมูลด้วย One-way ANOVA มีขนาดของกลุ่มตัวอย่างเท่ากัน ($n_1 = n_2$) และความแปรปรวนของคะแนนเท่ากัน ($\sigma_1^2 = \sigma_2^2$)

1. เมื่อต้องการเปรียบเทียบรายคู่ (Pair Comparison) และต้องการทราบค่าช่วงความเชื่อมั่น (Confidence Interval) ควรเลือกใช้ Tukey's A Test เพราะสามารถควบคุม α_{FW} ให้เท่ากับ α ที่กำหนดได้ รวมทั้งสถิตินี้มีอำนาจในการเปรียบเทียบสูงกว่าสถิติอื่นๆ

2. เมื่อต้องการเปรียบเทียบค่าเฉลี่ยของกลุ่มควบคุม (Control Group) กับกลุ่มทดลอง (Experimental Group) และต้องการทราบค่าช่วงความเชื่อมั่น ควรเลือกใช้ Dunnett Test เพราะสามารถควบคุม α_{FW} ให้เท่ากับ α ที่กำหนดได้

3. เมื่อต้องการเปรียบเทียบรายคู่ ควรเลือกใช้ REGW-Q Test เพราะค่า α_{FW} จะไม่มากกว่า α ที่กำหนด และสถิตินี้มีอำนาจ (power) ในการเปรียบเทียบสูงกว่า Tukey's A Test เมื่อไม่ต้องคำนวณหาช่วงความเชื่อมั่น

4. เมื่อต้องการเปรียบเทียบแบบซับซ้อน (Complex Comparison) ควรเลือกใช้ Scheffe Test เพราะสามารถควบคุมขนาดของ $\alpha_{EW} \leq 0.05$ ได้

2. เมื่อวิเคราะห์ข้อมูลด้วย One-way ANOVA มีขนาดของกลุ่มแตกต่างกัน ($n_1 \neq n_2$) แต่ความแปรปรวนของคะแนนเท่ากัน ($\sigma_1^2 = \sigma_2^2$)

1. เมื่อต้องการเปรียบเทียบรายคู่ ควรเลือกใช้ Tukey's B Test เพราะว่าจะระดับ α_{FW} จะไม่มากกว่า α ที่กำหนด และอาจจะน้อยกว่า

2. เมื่อต้องการเปรียบเทียบค่าเฉลี่ยของกลุ่มควบคุมกับของกลุ่มทดลองควรเลือกใช้ Dunnett Test เพราะว่าจะระดับ α_{FW} จะเท่ากับค่า α ที่กำหนดพอดี

3. ถ้าขนาดของกลุ่มตัวอย่างแตกต่างกันเล็กน้อย ควรเลือกใช้ Gabriel Test เพราะมีอำนาจในการเปรียบเทียบดีกว่าสถิติอื่น

4. ถ้าขนาดของกลุ่มตัวอย่างแตกต่างกันมาก ควรเลือกใช้ Hochberg's GT2 Test เพราะมีอำนาจในการเปรียบเทียบดีกว่าสถิติอื่น

3. เมื่อวิเคราะห์ข้อมูลด้วย ANOVA ตามรูปแบบการวัดซ้ำ (Repeated Measures)

1. เมื่อต้องการเปรียบเทียบรายคู่ ควรเลือกใช้ Tukey's A Test เพราะจะสามารถควบคุมระดับ α_{FW} ให้เท่ากับ α ที่กำหนดได้ เมื่อมีการกระทำซ้ำไม่มากครั้ง

2. อาจเลือกใช้ Paired-Samples t-test เมื่อต้องการเปรียบเทียบค่าเฉลี่ยของกลุ่มพลวิจัยจำนวนไม่มากครั้ง เพราะว่าหากเปรียบเทียบหลายครั้งจะเสี่ยงต่อการเกิด α_{FW} มาก (Howell, 2002)

4. เมื่อวิเคราะห์ข้อมูลด้วย One-way ANOVA และกลุ่มตัวอย่างมีความแปรปรวนของคะแนนแตกต่างกัน ($\delta_1^2 \neq \delta_2^2$)

1. เมื่อต้องการเปรียบเทียบรายคู่ ควรเลือกใช้ Dunnett's T3 Test หรือ Dunnett's C Test ก็ได้ เพราะว่าเป็นสถิติเชิงอนุรักษ์ (Conservative Test) กล่าวคือค่า α_{FW} จะน้อยกว่าค่า α ที่กำหนด

2. ถ้าขนาดของกลุ่มตัวอย่างมีน้อย ($n < 30$) ควรเลือกใช้ Dunnett's T3 Test

3. ถ้าขนาดของกลุ่มตัวอย่างมีมาก ($n > 30$) ใช้ Dunnett's C Test

5. เมื่อวิเคราะห์ข้อมูลด้วย k-way ANOVA และกลุ่มตัวอย่างมีขนาดเท่ากัน ($n_1 = n_2$)

1. ในกรณีที่ปฏิสัมพันธ์ร่วม (Interaction Effects) ของการวิจัยไม่มีนัยสำคัญ การทดสอบเพื่อเปรียบเทียบค่าเฉลี่ยของพลวิจัยในผลหลัก (Main Effects) แต่ละมิติหรือชุด (Family) สามารถใช้สถิติต่างๆที่กล่าวแล้วในข้อ ก. ข้างต้นได้

2. ปกติแล้ว ถ้าปฏิสัมพันธ์ร่วมของการวิจัยมีนัยสำคัญ นักวิจัยไม่ค่อยสนใจเปรียบเทียบค่าเฉลี่ยของพลวิจัยในผลหลัก แต่หากต้องการก็ควรเลือกใช้ Tukey's A Test เมื่อต้องการเปรียบเทียบรายคู่ และใช้ Dunnett Test เมื่อต้องการเปรียบเทียบค่าเฉลี่ยของกลุ่มควบคุมกับกลุ่มทดลอง เพราะว่าสถิติทั้งสองสามารถควบคุม α_{FW} ให้เท่ากับ α ที่กำหนดได้ [ทางเลือก (Option) นี้ไม่มีในโปรแกรม SPSS แต่มีใน SAS]

6. เมื่อวิเคราะห์ข้อมูลด้วย 2-way ANOVA ในรูปแบบผสม (Mixed Model) และกลุ่มตัวอย่างมีขนาดเท่ากัน ($n_1 = n_2$)

ในกรณีนี้ ควรเลือกใช้ Tukey's A Test หรือ Dunnett Test ก็ได้

7. เมื่อไม่ทราบข้อมูลใดๆเกี่ยวกับลักษณะค่าความแปรปรวนของคะแนนของกลุ่มพลวิจัย

1. ในกรณีนี้ ผู้วิจัยควรเลือกใช้ Bonferroni Test หากต้องการทราบค่าช่วง

ความเชื่อมั่น เพราะวาสถิตินี้สามารถควบคุมค่า $\alpha_{FW} \leq \alpha$ ได้

2. ในกรณีสงสัยว่าค่าความแปรปรวนของพลวิจัยเท่ากันหรือไม่ ควรเลือกใช้ Games-Howell Test เพราะวาสถิตินี้สามารถใช้ได้กับข้อมูลที่มีค่าความแปรปรวนของคะแนนของกลุ่มพลวิจัยเท่ากัน ($\delta_1^2 = \delta_2^2$) หรือแตกต่างกัน ($\delta_1^2 \neq \delta_2^2$) ก็ได้

สรุป

เนื่องจากสถิติที่ใช้เพื่อการทดสอบว่าค่าเฉลี่ยของกลุ่มพลวิจัยแตกต่างกันหรือไม่มากมาย (ประมาณ 20 อย่าง) แต่ละอย่างมีข้อดีและข้อเสียแตกต่างกัน และเหมาะสมสำหรับข้อมูลและวัตถุประสงค์ของการวิจัยที่แตกต่างกัน รวมทั้งสามารถทำให้เกิดอัตราความคลาดเคลื่อนในการทดสอบได้แตกต่างกัน ดังนั้น นักวิจัยควรที่จะเลือกใช้สถิติต่างๆ เหล่านี้ให้สอดคล้องกับลักษณะข้อมูลและวัตถุประสงค์ของตน ทั้งนี้เพื่อให้ผลการวิจัยมีความถูกต้องมากที่สุด ซึ่งเป็นสิ่งที่บรรดา

นักวิจัยทั้งหลายปรารถนา แต่หากว่าท่านไม่สะดวกที่จะจดจำรายละเอียดของการใช้สถิติต่างๆ เหล่านี้ ก็ควรที่จะศึกษาคำแนะนำทั่วไปในการใช้สถิติเพื่อการเปรียบเทียบข้างต้น และสิ่งที่สำคัญที่สุดสำหรับนักวิจัยทุกท่านก็คือ ท่านควรอย่างยิ่งที่จะตอบคำถามที่ผู้อื่นถามท่านว่า ท่านเลือกใช้สถิติบางอย่างในงานวิจัยของท่านเพราะเหตุใด และคำตอบที่ท่านควรจะตอบก็ไม่ควรจะเป็น "เพราะคนอื่นก็ใช้จึงใช้ตาม" เพราะนั่นแสดงว่า ท่านไม่เข้าใจการเลือกใช้สถิติให้เหมาะสมกับวัตถุประสงค์ของการวิจัย และธรรมชาติของข้อมูลของท่านเลย แล้วท่านจะเข้าใจวิธีการแปลผลความหมายการวิจัยของท่านอย่างลึกซึ้งหรือถูกต้องได้อย่างไร

อนึ่ง สำหรับท่านที่สนใจสูตรในการคำนวณของสถิติแต่ละอย่างที่ผู้เขียนได้กล่าวถึงในบทความนี้ โปรดหาดูได้ที่ Appendix 10: Post Hoc Tests ในคู่มือของ SPSS ตั้งแต่ Version 7.0 เป็นต้นมา หรือที่ SPSS 14.0. (2008) ก็ได้

รองศาสตราจารย์ ดร. สุวัฒน์ สุกมลสันต์

รองศาสตราจารย์ ดร. สุวัฒน์ สุกมลสันต์ วุฒิปริญญาโท (เกียรตินิยม), ค.ม. (การสอนภาษาอังกฤษ), M.Sc (Applied Linguistics) และ ค.ด. (การวัดและประเมินผลทางการศึกษา) เคยสอนภาษาอังกฤษ สถิติเพื่อการวิจัย วิทยาการวิจัย การวัดและประเมินผลทางภาษาและการศึกษาที่สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ปัจจุบันปฏิบัติงานที่มหาวิทยาลัยแม่โจ้

เอกสารอ้างอิง

- Atzenbeck, C. (2010). **Tests for Group Comparisons With More Than 3 Groups**. Retrieved from <http://cs.aau.dk/~claus/temp/TestOverview.pdf/> on May 10, 2010.
- Becker, L.A. (1999). **One-way: Single-Factor, Between-Subjects Designs**. University of Colorado - Colorado Springs. Retrieved from <http://web.uccs.edu/lbecker/SPSS80/oneway.htm/> on May 10, 2010.
- Cabral, H.J. (2008). **Multiple Comparisons Procedures**. Retrieved from <http://circ.ahajournals.org/cgi/content/full/117/5/698/> on May 10, 2010.
- ESF. (2010). **Multiple Comparison Tests**. Retrieved from http://www.esf.edu/for/bevilacqua/apm620/multiple_comparisons/sld001.htm/ on May 10, 2010.
- Field, A. (2000). **Contrast and Post Hoc Tests for One-way ANOVA**. Retrieved from <http://www.statisticshell.com/contrasts.pdf/> on May 10, 2010.
- FSU. (2010). **Dunn-Sidak Test**. Retrieved from http://mailer.fsu.edu/~akamata/edf5402/05_Mult%20Comp_3.pdf/ on May 10, 2010.
- Garson, G.D. (2009). **ANOVA and ANCOVA**. Retrieved from <http://faculty.chass.ncsu.edu/garson/PA765/anova.htm/> on May 10, 2010.
- Gold. (2010). **Multiple Comparisons**. Retrieved from <http://homepages.gold.ac.uk/aphome/lec13ohp.ppt/> on May 10, 2010.
- GraphPad. (2010). **What are Planned Comparisons?**. Retrieved from <http://www.graphpad.com/faq/viewfaq.cfm?faq=1091/> on May 10, 2010.
- Howell, D.C. (2002). **Multiple Comparisons with Repeated Measures**. Retrieved from http://www.uvm.edu/~dhowell/StatPages/More_Stuff/RepMeasMultComp/RepMeasMultComp.html/ on May 10, 2010.
- Karpinski, A. (2006). **Planned Comparisons and Post Hoc Tests**. Retrieved from http://astro.temple.edu/~andykarp/psych522524/06_contrasts2.pdf/ on May 10, 2010.
- Maxwell and Delaney. (2010). **Multiple Comparisons**. Retrieved from http://www.psychology.mcmaster.ca/bennett/psy710/lectures/maxwell_chp5.pdf/ on May 10, 2010.
- Neel, J.H. (1997). **An Introduction to Multiple Comparison Techniques**. Retrieved from

- <http://education.gsu.edu/jneel/courses/eprs8540/handouts/multcomp.pdf/> on May 10, 2010.
- Newsom, (2009). **Post Hoc Tests**. Retrieved from http://www.upa.pdx.edu/IOA/newsom/da1/ho_posthoc.doc/ on May 10, 2010.
- Pierce, T. (2003). **Comparisons Between Treatment Means in ANOVA**. Department of Psychology, Radford University. Retrieved from on May 10, 2010.
- Prism Guide. (2010). **Choosing One-Way ANOVA and Related Analyses**. Retrieved from http://www.graphpad.com/articles/interpret/anova/choosing_test.htm/ on May 10, 2010.
- SPSS 14.0. (2008). **Appendix 10: Post Hoc Tests**. Retrieved from http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/algorithms/14.0/app10_post_hoc_tests.pdf/ on May 10, 2010.
- SPSS 17.0. (2010). **GLM Multivariate Post Hoc Comparisons**. Retrieved from <http://support.spss.com/ProductsExt/SPSS/Documentation/Manuals/17.0/SPSS%20Advanced%20Models%2015.0.pdf/> on May 10, 2010.
- SSTARS. (2009). **Pairwise Comparisons in SAS and SPSS**. Retrieved from http://www.uky.edu/ComputingCenter/SSTARS/www/documentation/MultipleComparisons_3.htm/ on May 10, 2010.
- Stevens, (1999). **Post Hoc Tests in ANOVA**. Retrieved from <http://www.uoregon.edu/~stevensj/posthoc.pdf/> on May 10, 2010.
- Wikipedia. (2010). **Multiple Comparisons**. Retrieved from http://en.wikipedia.org/wiki/Multiple_comparisons/ on May 10, 2010.
- Willson, V. (2010). **Hypotheses About Contrasts**. Retrieved from [http://www.coe.tamu.edu/~vwillson/Videos/epsy641/LEC10641/LEC10.ppt#262,1,Hypotheses about Contrasts/](http://www.coe.tamu.edu/~vwillson/Videos/epsy641/LEC10641/LEC10.ppt#262,1,Hypotheses%20about%20Contrasts/) on May 10, 2010.
