

สิ่งสำคัญที่ควรคำนึงในการเปรียบเทียบรายคู่

พอใจ พัทธนิตย์ธรรม

ภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

บทคัดย่อ

การวิเคราะห์ความแปรปรวนเป็นวิธีการทางสถิติสำหรับเปรียบเทียบค่าเฉลี่ยประชากรมากกว่าสองกลุ่มขึ้นไป เมื่อผลการทดสอบพบว่ามีความแตกต่างอย่างมีนัยสำคัญ การเปรียบเทียบรายคู่ (multiple comparisons) จะถูกนำมาใช้เพื่อระบุว่าค่าเฉลี่ยประชากรคู่ใดมีความแตกต่างกัน บทความนี้นำเสนอแนวคิดของการเปรียบเทียบรายคู่แบบขั้นตอนเดียวด้วยวิธีการของ Bonferroni Scheffe และ Turkey วิธีการของ Bonferroni มีประสิทธิภาพดีที่สุดสำหรับการเปรียบเทียบเพียงบางคู่ ขณะที่วิธีการที่เป็นที่ยอมรับมากที่สุดสำหรับการเปรียบเทียบทุกคู่คือวิธีการของ Turkey การพิจารณาเฉพาะค่า p (p-values) ที่ได้จากการเปรียบเทียบรายคู่ไม่เพียงพอสำหรับการแปลผลของการเปรียบเทียบรายคู่ ช่วงเชื่อมั่นค่าเฉลี่ยของแต่ละกลุ่มที่นำมาเปรียบเทียบ และช่วงเชื่อมั่นผลต่างค่าเฉลี่ยประชากรของคู่ที่เปรียบเทียบ ควรนำมาสรุปผลร่วมกับความแตกต่างที่มีความหมายต่อการใช้งาน จะให้ข้อเสนอแนะที่สามารถนำผลการศึกษามาตัดสินใจใช้เปลี่ยนแปลงการปฏิบัติงานได้

คำสำคัญ: สิ่งสำคัญที่ควรคำนึงถึง, การเปรียบเทียบรายคู่

การเปรียบเทียบรายคู่ (multiple comparisons) เป็นวิธีการทดสอบความแตกต่างของค่าเฉลี่ยประชากรระหว่างสองกลุ่มใด ๆ ในกรณีที่พบว่าผลการวิเคราะห์ความแปรปรวนทางเดียวมีนัยสำคัญทางสถิติ ปัจจุบันการเปรียบเทียบรายคู่มีหลากหลายวิธีให้นักวิจัยเลือกใช้ เช่น วิธีการของ bonferroni scheffe และ tukey การเปรียบเทียบรายคู่แต่ละวิธีมีความไวของการตรวจพบความแตกต่างค่าเฉลี่ยประชากรที่แตกต่างกันไม่มากนัก แต่สิ่งสำคัญที่ทำให้การสรุปผลแตกต่างกัน คือ การกำหนดจำนวนคู่เปรียบเทียบ นอกจากนี้หากนักวิจัยพิจารณาถึงประโยชน์ของการนำมาใช้งานร่วมในการสรุปผลที่ได้จากการเปรียบเทียบรายคู่จะทำให้การสรุปผลการวิจัยมีความเหมาะสมยิ่งขึ้น

ในบทความนี้เป็นการสรุปแนวคิด หลักการของการเปรียบเทียบรายคู่ พร้อมทั้งแนวทางในการแปลผลที่ได้จากการเปรียบเทียบรายคู่

จำนวนคู่กับโอกาสในการพบความแตกต่าง

เมื่อผลการเปรียบเทียบค่าเฉลี่ยประชากรมากกว่า 2 กลุ่มที่อิสระกันจากการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA) พบว่า ปฏิเสธสมมติฐานหลัก ทำให้นักวิจัยทราบเพียงว่ามีค่าเฉลี่ยประชากรอย่างน้อย 1 คู่ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่ไม่สามารถระบุได้ว่าค่าเฉลี่ยประชากรคู่ใดแตกต่างกัน ดังนั้นถ้าต้องการทราบว่ามีความแตกต่างของค่าเฉลี่ยคู่หนึ่งคู่ใด นักวิจัยจะต้องทำการทดสอบเปรียบเทียบรายคู่ต่อไป

ในการเปรียบเทียบรายคู่ทีละคู่โดยใช้สถิติที (Independent t-test) จะไม่เหมาะสม ทั้งนี้เพราะความผิดพลาดแบบที่ 1 (Type I error) ในการสรุปผลจะเพิ่มขึ้นตามจำนวนคู่เปรียบเทียบที่เพิ่มขึ้น⁽¹⁻⁵⁾ เช่น นักวิจัยกำหนดระดับนัยสำคัญ (α) ไว้เท่ากับ 0.05 เมื่อทดสอบคู่ที่ 1 โอกาสสรุปผิดเท่ากับ 0.05 และโอกาส

สรุปถูกเท่ากับ 0.95 เมื่อทดสอบคู่ที่ 2 โอกาสสรุปผิดเท่ากับ 0.05 และโอกาสสรุปถูกเท่ากับ 0.95 เช่นกัน ในการสรุปผลการเปรียบเทียบรายคู่จำนวน 2 คู่ได้ถูกต้องจะต้องรวมโอกาสสรุปถูกต้องทั้งสองครั้งเข้าด้วยกันซึ่งจะเท่ากับ $0.95 \times 0.95 = 0.9025$ และโอกาสสรุปผิดจะเท่ากับ $1 - 0.9025 = 0.0975$ จะเห็นได้ว่าโอกาสสรุปผิดมีค่าเพิ่มขึ้นจาก 0.05 เป็น 0.0975 ดังนั้น ยังมีจำนวนคู่เปรียบเทียบมากคู่เท่าไร โอกาสที่จะสรุปผลผิดจะยิ่งมากขึ้น

การเปรียบเทียบรายคู่

เพื่อแก้ไขปัญหาความผิดพลาดที่เพิ่มขึ้นในการเปรียบเทียบรายคู่ นักสถิติจึงพัฒนาวิธีการเปรียบเทียบรายคู่ขึ้นหลายวิธี การเปรียบเทียบรายคู่สามารถจำแนกออกเป็น 2 ประเภท

ประเภทที่ 1 แบบขั้นตอนเดียว (Single-step procedure) การทดสอบแบบนี้จะกำหนดโอกาสสรุปผิดของการเปรียบเทียบแต่ละคู่เท่ากัน และมีการคำนวณโอกาสสรุปผิดเพียงครั้งเดียว เช่น วิธีการของ Bonferroni Scheffe และ Tukey⁽¹⁾

ประเภทที่ 2 แบบทีละขั้น (Stepwise procedure) จะกำหนดให้โอกาสสรุปผิดของการเปรียบเทียบแต่ละคู่แตกต่างกัน และมีการคำนวณโอกาสสรุปผิดทุกครั้งที่ของการเปรียบเทียบคู่ใด ๆ เช่น วิธีการของ Holm Hochberg's และ Hommel's⁽¹⁾

วิธีการทดสอบเปรียบเทียบรายคู่ทั้งแบบขั้นตอนเดียวและแบบทีละขั้นจะควบคุมโอกาสสรุปผิดทั้งหมดให้มีค่าไม่เกินระดับนัยสำคัญที่กำหนดไว้

บทความนี้จะไม่กล่าวถึงวิธีการเปรียบเทียบรายคู่แบบทีละขั้นเพราะมีวิธีการคำนวณที่ซับซ้อน แต่จะกล่าวถึงการเลือกใช้การเปรียบเทียบรายคู่แบบขั้นตอนเดียววิธีการของ Bonferroni Scheffe และ Tukey เพราะมีวิธีการคำนวณที่เข้าใจง่ายและเป็นที่ยอมรับใช้กันทั่วไป

วิธีการเปรียบเทียบรายคู่แบบขั้นตอนเดียวแบ่งออกได้เป็น 2 กลุ่ม คือ ระดับนัยสำคัญแต่ละคู่เปรียบเทียบขึ้นอยู่กับจำนวนคู่ และ ระดับนัยสำคัญแต่ละคู่

เปรียบเทียบไม่ขึ้นกับจำนวนคู่

การเปรียบเทียบรายคู่แบบขั้นเดียว กรณีระดับนัยสำคัญแต่ละคู่เปรียบเทียบขึ้นอยู่กับจำนวนคู่ ได้แก่ วิธีการของ Bonferroni วิธีนี้พัฒนามาจากสถิติที่ สถิติทดสอบของ Bonferroni มีสูตรในการคำนวณเหมือนกับสถิติที่ แต่ใช้ค่าส่วนเบี่ยงเบนกำลังสองเฉลี่ยภายในกลุ่ม (within mean square) ที่ได้จากการวิเคราะห์ความแปรปรวนมาเป็นค่าความแปรปรวนร่วม (pooled variance) และใช้ค่าระดับความเป็นอิสระ (degree of freedom) ของสถิติที่เท่ากับระดับความเป็นอิสระของส่วนเบี่ยงเบนกำลังสองเฉลี่ยภายในกลุ่มที่ได้จากการวิเคราะห์ความแปรปรวน การหาค่าวิกฤตวิธีการของ Bonferroni เป็นค่าที่ได้จากการที่แต่ใช้ระดับนัยสำคัญของการทดสอบแต่ละคู่เท่ากับ α/k โดยที่ k คือ จำนวนคู่เปรียบเทียบ⁽¹⁻⁴⁾

ในกรณีที่ตั้งระดับนัยสำคัญของการทดสอบไว้เท่ากับ 0.05 ถ้าเปรียบเทียบ 3 คู่ ระดับนัยสำคัญของการเปรียบเทียบแต่ละคู่เท่ากับ $0.05/3 = 0.0167$ ถ้าเปรียบเทียบ 4 คู่ ระดับนัยสำคัญของการเปรียบเทียบแต่ละคู่เท่ากับ $0.05/4 = 0.0125$ การเปรียบเทียบรายคู่ด้วยวิธีนี้ถ้ามีจำนวนคู่เปรียบเทียบมากโอกาสที่จะพบความแตกต่างจะยิ่งน้อย

การเปรียบเทียบรายคู่แบบขั้นเดียว กรณีระดับนัยสำคัญแต่ละคู่เปรียบเทียบไม่ขึ้นกับจำนวนคู่ ได้แก่ วิธีการของ Scheffe และ Tukey

วิธีการของ Scheffe มีวิธีการคำนวณสถิติ S ในการเปรียบเทียบรายคู่โดยมีสูตรคำนวณ⁽⁶⁾ ดังนี้

$$S = \frac{|\bar{x}_1 - \bar{x}_2|}{SE} \quad \text{โดยที่ } SE = \sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

ค่า p-value ที่ได้จากสถิติ S สามารถนำไปเปรียบเทียบกับระดับนัยสำคัญที่ตั้งไว้ได้โดยตรงโดยไม่ต้องปรับค่าระดับนัยสำคัญตามจำนวนคู่ที่จะเปรียบเทียบ

วิธีการของ Tukey มีวิธีการคำนวณสถิติ q ในการเปรียบเทียบรายคู่โดยมีสูตรคำนวณ⁽⁶⁾ ดังนี้

สิ่งสำคัญที่ควรคำนึงในการเปรียบเทียบรายคู่

$$q = \frac{|\bar{x}_1 - \bar{x}_2|}{SE.} \quad \text{โดยที่ } SE. = \sqrt{\frac{S^2}{n}}$$

ค่า p-value ที่ได้จากสถิติ q สามารถนำไปเปรียบเทียบระดับนัยสำคัญได้โดยตรงเช่นเดียวกับวิธีการของ Scheffe

การเลือกใช้วิธีการเปรียบเทียบรายคู่

หากนักวิจัยมีวัตถุประสงค์ที่ชัดเจนในการเปรียบเทียบรายคู่เพียงบางคู่ ควรเลือกใช้วิธีการของ Bonferroni เพราะทำให้มีโอกาสที่จะพบความแตกต่างของค่าเฉลี่ยประชากรเพิ่มขึ้น (พิจารณาตัวอย่างที่ 1)

ตัวอย่างที่ 1 การศึกษาผลการกำจัดลูกน้ำยุงลายด้วยการใช้ผลิตภัณฑ์เคมีทรายกำจัดลูกน้ำ (ทรายอะเบท) สารสกัดชีวภาพที่หมักด้วยสะเดา และกลุ่มควบคุมซึ่งเป็นน้ำฝน การทดลองนี้ทำในห้องปฏิบัติการโดยเปรียบเทียบจำนวนลูกน้ำยุงลายที่เหลืออยู่ในจานทดลองเมื่อเวลาผ่านไป 24 ชั่วโมง (จานทดลองแต่ละจานมีลูกน้ำยุงลายจำนวน 100 ตัวเมื่อเริ่มต้นการทดลอง แต่ละวิธี

มีจานทดลองจำนวน 10 จาน)

นักวิจัยมีวัตถุประสงค์เพื่อเปรียบเทียบจำนวนลูกน้ำยุงลายที่เหลืออยู่ระหว่างการใส่สารสกัดสะเดากับทรายอะเบท เพราะต้องการค้นหาวิธีที่ปลอดภัยในการกำจัดลูกน้ำยุงลายแทนการใช้สารเคมี และเปรียบเทียบจำนวนลูกน้ำยุงลายที่เหลืออยู่ระหว่างการใช้น้ำฝนกับสารสกัดสะเดา เนื่องจากการใช้สารสกัดสะเดาจะมีประโยชน์ต่อการนำไปใช้ก็ต่อเมื่อมีประสิทธิภาพดีกว่ากลุ่มควบคุม โดยไม่สนใจผลการเปรียบเทียบของกลุ่มควบคุมกับทรายอะเบท เพราะเป็นที่ทราบแล้วว่าทรายอะเบทมีประสิทธิภาพในการกำจัดลูกน้ำยุงลายจริง

จากผลการวิเคราะห์ความแปรปรวนทางเดียวในตารางที่ 1 ค่าเฉลี่ยจำนวนลูกน้ำยุงลายที่เหลืออยู่เมื่อเวลาผ่านไป 24 ชั่วโมงด้วยการใช้ทรายอะเบท สารสกัดสะเดา และน้ำฝน มีความแตกต่างกันอย่างมีนัยสำคัญ แสดงว่า มีอย่างน้อย 1 วิธีที่แตกต่างกัน นักวิจัยจึงทำการเปรียบเทียบรายคู่ต่อ ผลของการเปรียบเทียบรายคู่ด้วยวิธีการของ Bonferroni ในตารางที่ 2 กรณีที่ทำการเปรียบเทียบทั้ง 3 วิธี จำนวน 3 คู่ ไม่พบความ

ตารางที่ 1 ผลลัพธ์การวิเคราะห์ความแปรปรวนทางเดียวจากโปรแกรมทางสถิติ

แหล่งความแปรปรวน	ระดับความเป็นอิสระ	ผลรวมส่วนเบี่ยงเบนกำลังสอง	ส่วนเบี่ยงเบนกำลังสองเฉลี่ย	ค่าสถิติเอฟ	ค่า p-value
ระหว่างกลุ่ม	2	56397	25198.8	22986	<0.001
ภายในกลุ่ม	27	33	1.23		
รวม	29	56430			

ตารางที่ 2 ค่าระดับนัยสำคัญของการเปรียบเทียบแต่ละคู่ (α_i) และ p-value ของการเปรียบเทียบรายคู่ด้วยวิธีของ Bonferroni

จำนวนคู่ และวิธีการ	คู่เปรียบเทียบ	ค่า α_i	ค่า p-value	ผลการทดสอบ
เปรียบเทียบจำนวนคู่ วิธี Bonferroni				
1 คู่	สารสกัดสะเดา กับ ทรายอะเบท	0.05	0.023	มีนัยสำคัญ
2 คู่	สารสกัดสะเดา กับ ทรายอะเบท	0.025	0.023	มีนัยสำคัญ
3 คู่	สารสกัดสะเดา กับ ทรายอะเบท	0.0167	0.023	ไม่มีนัยสำคัญ

แตกต่างกันจำนวนลูกน้ำยุงลายที่เหลืออยู่ของการใช้สารสกัดสะเดากับทรายอะเบท (p -value=0.069) แต่หากเลือกเปรียบเทียบเฉพาะคู่ที่สนใจ คือ สารสกัดสะเดากับทรายอะเบท และ นำฝนกับสารสกัดสะเดา จำนวน 2 คู่ จะพบความแตกต่างจำนวนลูกน้ำยุงลายที่เหลืออยู่ของการใช้ สารสกัดสะเดากับทรายอะเบท (p -value =0.046) จะเห็นได้ว่าเมื่อจำนวนคู่ของการทดสอบลดลงจาก 3 คู่เป็น 2 คู่ ผลการเปรียบเทียบสารสกัดสะเดากับทรายอะเบทจะพบความแตกต่างอย่างมีนัยสำคัญ

การเปรียบเทียบด้วยวิธีการของ Bonferroni ปัจจัยสำคัญที่ทำให้โอกาสพบความแตกต่างมีมากหรือน้อยคือจำนวนคู่เปรียบเทียบ เมื่อจำนวนคู่ลดลง ค่าระดับนัยสำคัญของการเปรียบเทียบแต่ละคู่จะลดลงตามจำนวนคู่ ทำให้มีโอกาสที่จะพบความแตกต่างลดลงเมื่อจำนวนคู่เพิ่มขึ้น ในตารางที่ 2 เมื่อจำนวนคู่ เท่ากับ 3 ค่า p -value เป็น 0.023 นำค่า p -value นี้เปรียบเทียบกับระดับนัยสำคัญ 0.0167 ทำให้ไม่พบความแตกต่าง เมื่อจำนวนคู่ลดเหลือ 2 และ 1 คู่ นำค่า p -value ที่ได้เปรียบเทียบกับระดับนัยสำคัญ 0.025 และ 0.05 ตามลำดับ ผลการทดสอบจะพบความแตกต่าง ดังนั้นวิธีการของ Bonferroni ถ้าจำนวนคู่น้อยจะมีโอกาสพบความแตกต่างได้มาก

การเปรียบเทียบด้วยวิธีการของ Bonferroni Scheffe และ Tukey เมื่อเปรียบเทียบทุกคู่ (จำนวนคู่ 3

คู่) ในตารางที่ 3 ให้ผลการทดสอบเหมือนกันว่า ไม่พบความแตกต่าง แต่เมื่อพิจารณาค่า p -value ที่ได้จะพบว่าทั้ง 3 วิธีดังกล่าวมีโอกาสที่จะพบความแตกต่าง หรือมีความไวในการทดสอบที่แตกต่างกัน โดยวิธีการของ Tukey มีโอกาสพบความแตกต่างมากที่สุด เพราะมีค่า p -value (0.058) น้อยที่สุด ดังนั้น นักวิจัยควรเลือกใช้วิธีการของ Tukey เพื่อระบุความแตกต่างระหว่างค่าเฉลี่ยประชากรคู่ใด ๆ แล้วนำประโยชน์ในการใช้งานมาพิจารณาคู่ควบกันดังจะกล่าวในหัวข้อถัดไป

แนวทางในการแปลผลที่ได้จากการเปรียบเทียบรายคู่

การเปรียบเทียบรายคู่มีหลายวิธี การเลือกใช้วิธีใดจะขึ้นอยู่กับเงื่อนไขและข้อตกลงเบื้องต้นของวิธีการทางสถิติในแต่ละวิธี และความนิยมใช้ในแต่ละกลุ่มวิชาชีพ ค่าสถิติที่คำนวณได้จากแต่ละวิธีส่วนใหญ่ให้ข้อสรุปที่ไม่แตกต่างกัน แต่การสรุปผลที่ไม่เหมาะสมส่วนใหญ่อุบัติจากการที่นักวิจัยไม่ได้นำประโยชน์ของการใช้งานมาพิจารณาร่วมด้วย พิจารณาตัวอย่างที่ 2 การเปรียบเทียบรายคู่ด้วยวิธีการของ Bonferroni Scheffe และ Tukey ซึ่งวิธีดังกล่าวเป็นที่นิยมใช้กันทั่วไป พร้อมแนวทางในการพิจารณาประโยชน์ของการใช้งานร่วมกับผลทางทดสอบทางสถิติ

ตัวอย่างที่ 2 ทีมนักวิจัยจัดโครงการให้ความรู้และทักษะการป้องกันไข้หวัดนก โดยให้เจ้าหน้าที่สาธารณสุขเป็นผู้อบรมให้แก่อาสาสมัครสาธารณสุข

ตารางที่ 3 ค่า p -value ของการเปรียบเทียบรายคู่ด้วยวิธีของ Bonferroni Scheffe และ Tukey เมื่อจำนวนคู่เปรียบเทียบเท่ากับ 3 คู่

จำนวนคู่ และวิธีการ	คู่เปรียบเทียบ	ค่า α	ค่า p -value	ผลการทดสอบ
เปรียบเทียบระหว่างวิธีในการทดสอบ 3 คู่				
Tukey	สารสกัดสะเดา กับ ทรายอะเบท	0.05	0.058	ไม่มีนัยสำคัญ
Bonferroni	สารสกัดสะเดา กับ ทรายอะเบท	$(0.0167 \times 3) = 0.05$	$(0.023 \times 3) = 0.069$	ไม่มีนัยสำคัญ
Scheffe	สารสกัดสะเดา กับ ทรายอะเบท	0.05	0.072	ไม่มีนัยสำคัญ

สิ่งสำคัญที่ควรคำนึงในการเปรียบเทียบรายคู่

(อสม.) ผู้นำครอบครัวที่เป็นหญิง (แม่บ้าน) และผู้นำครอบครัวที่เป็นชาย (พ่อบ้าน) โดยโครงการดังกล่าวนี้จะบรรลุเป้าหมายก็ต่อเมื่อประชาชนทั้งสามกลุ่มดังกล่าวมีคะแนนความรู้ไม่น้อยกว่า 80 คะแนน (คะแนนเต็ม 100) คะแนนความรู้และทักษะในการป้องกันไข้

ตารางที่ 4 คะแนนความรู้และทักษะการป้องกันไข้หวัดนกของตัวอย่าง อสม. แม่บ้าน และ พ่อบ้าน

กลุ่ม	จำนวน (คน)	\bar{x} (SD.)	95% CI.
อสม.	50	85 (4.7)	84 ถึง 86
แม่บ้าน	50	81 (4.2)	80 ถึง 82
พ่อบ้าน	50	83 (4.5)	82 ถึง 84

หวัดนกแสดงในตารางที่ 4

จากตารางที่ 6 ผลลัพธ์จากการเปรียบเทียบรายคู่ด้วยวิธีการของ Bonferroni Scheffe และ Tukey ให้ผลการทดสอบทางสถิติไปในทิศทางเดียวกัน คือพบความแตกต่างคะแนนความรู้และทักษะการป้องกันไข้หวัดนกของ อสม.กับแม่บ้าน และ อสม.กับพ่อบ้าน หากนักวิจัยพิจารณาสรุปผลจากการทดสอบทางสถิติอย่างเดียว อาจจะสรุปว่าแม่บ้านและพ่อบ้านมีความรู้และทักษะการป้องกันไข้หวัดนกลดกว่า อสม. และเสนอให้มีการอบรมพ่อบ้าน และแม่บ้านให้มีความรู้เพิ่มขึ้นเท่ากับ อสม.

แต่ถ้านักวิจัยพิจารณาช่วงเชื่อมั่นคะแนนของแม่บ้านและพ่อบ้านที่มีค่าระหว่าง 80 ถึง 82 คะแนน และ 82 ถึง 84 คะแนน ตามลำดับ จะพบว่าแม่บ้าน

ตารางที่ 5 ผลลัพธ์การวิเคราะห์ความแปรปรวนทางเดียวจากโปรแกรมทางสถิติ

แหล่งความแปรปรวน	ระดับความเป็นอิสระ	ผลรวมส่วนเบี่ยงเบนกำลังสอง	ส่วนเบี่ยงเบนกำลังสองเฉลี่ย	ค่าสถิติเอฟ	ค่า p-value
ระหว่างกลุ่ม	2	402.1	201	10.1	<0.001
ภายในกลุ่ม	147	2981.9	19.8		
รวม	149	3321			

ตารางที่ 6 ผลลัพธ์จากโปรแกรมทางสถิติของการเปรียบเทียบรายคู่ด้วยวิธีของ Tukey, Bonferroni และ Scheffe

คู่เปรียบเทียบ	วิธีการ	p-value	95% CI
อสม. กับ แม่บ้าน	Tukey	<0.001	1.91 ถึง 6.09
	Bonferroni	<0.001	1.84 ถึง 6.15
	Scheffe	<0.001	1.79 ถึง 6.20
อสม. กับ พ่อบ้าน	Tukey	0.028	0.20 ถึง 4.38
	Bonferroni	0.034	0.13 ถึง 4.44
	Scheffe	0.040	0.08 ถึง 4.49
แม่บ้าน กับ พ่อบ้าน	Tukey	0.133	-3.80 ถึง 0.38
	Bonferroni	0.171	-3.87 ถึง 0.45
	Scheffe	0.162	-3.91 ถึง 0.49

และพ่อบ้านมีความรู้เพียงพอแล้ว สรุปได้ว่าโครงการนี้บรรลุเป้าหมายที่กำหนดไว้เพราะชุมชนนี้มี อสม. แม่บ้าน และ พ่อบ้าน ที่มีความรู้และทักษะการป้องกันไข้หวัดนกอยู่ในระดับที่ดีเพียงพอแล้วทั้งสามกลุ่ม ถึงแม้ อสม. จะรู้มากกว่า แต่เป็นความแตกต่างภายในกลุ่มที่มีความรู้เพียงพอ จึงไม่มีความจำเป็นจะต้องมีโครงการเสริมความรู้และทักษะในการป้องกันไข้หวัดนกเพิ่มเติม

การเปรียบเทียบรายคู่ในโปรแกรมทางสถิติ

ปัจจุบันมีโปรแกรมทางสถิติหลายโปรแกรมให้นักวิจัยเลือกใช้เมื่อต้องการเปรียบเทียบรายคู่ ผลลัพธ์ที่ได้จากโปรแกรมส่วนใหญ่ เช่น SPSS และ STATA จะรายงานผลต่างค่าเฉลี่ยของแต่ละคู่เปรียบเทียบ พร้อมค่า p-value ที่ได้จากการทดสอบทั้งหมดทุกคู่ ดังนั้นหากนักวิจัยต้องการเปรียบเทียบทุกคู่ ด้วยสถิติ Scheffe หรือ Tukey ก็สามารถนำค่า p-value นี้ไปเปรียบเทียบโดยตรงกับระดับนัยสำคัญที่ตั้งไว้

สำหรับการรายงานค่า p-value ของวิธี Bonferroni โปรแกรมจะนำค่า p-value ที่คำนวณได้คูณด้วยจำนวนคู่เพื่อให้นักวิจัยมีความสะดวกในการนำค่า p-value นี้ไปเปรียบเทียบกับระดับนัยสำคัญที่ตั้งไว้ได้โดยตรง (ปรับค่า p-value แทนปรับค่า α) หากนักวิจัยต้องการเปรียบเทียบเฉพาะคู่ที่สนใจบางคู่ จะต้องนำค่า p-value ที่โปรแกรมคำนวณให้ มาหารด้วยจำนวนคู่ทั้งหมดเพื่อให้ได้ค่า p-value แท้จริงของคู่เปรียบเทียบนั้น ๆ แล้วนำไปเปรียบเทียบกับระดับนัยสำคัญที่ปรับตามจำนวนคู่ที่ต้องการเปรียบเทียบ เช่น ในตารางที่ 3 ผลลัพธ์จากโปรแกรมรายงานค่า p-value ที่ได้จากการเปรียบเทียบจำนวนคู่ 3 คู่ เท่ากับ 0.069 สมมตินักวิจัยต้องการเปรียบเทียบเพียง 2 คู่ ดังนั้น ค่า p-value ที่แท้จริงคือ $0.069/3 = 0.023$ นำค่า p-value นี้ไปเปรียบเทียบกับ 0.025 ($\alpha/2 = 0.05/2 = 0.025$)

บทสรุป

การเปรียบเทียบรายคู่มีหลายวิธี นักวิจัยควรเลือก

ใช้วิธีที่เหมาะสมกับข้อตกลงเบื้องต้นของการใช้สถิติ นั้น ๆ วิธีการของ Bonferroni เป็นตัวเลือกที่ดีเมื่อนักวิจัยต้องการเปรียบเทียบรายคู่เฉพาะคู่ที่สนใจ จำนวน 1 หรือ 2 คู่ หากต้องการเปรียบเทียบจำนวนคู่มากกว่านี้ควรใช้วิธีของ Scheffe หรือ Tukey

ถ้าพิจารณาจากความไวของการพบความแตกต่างวิธีของ Tukey จะมีความไวในการพบคู่ที่แตกต่างดีกว่าวิธีของ Bonferroni และ Scheffe

การแปลผลการเปรียบเทียบรายคู่ที่พบในงานวิจัยส่วนใหญ่ไม่ได้นำขนาดความแตกต่างที่มีประโยชน์ของการใช้งานมาพิจารณาร่วมด้วย ทำให้ข้อเสนอแนะที่ได้จากงานวิจัยอาจไม่เหมาะสมต่อการนำไปปฏิบัติ

ดังนั้นในกรณีที่นักวิจัยมีการพิจารณาขนาดความแตกต่างที่มีประโยชน์ของการใช้งานร่วมกับผลการเปรียบเทียบรายคู่ การเลือกใช้สถิติที่มีความไวในการระบุความแตกต่างจะเป็นทางเลือกที่ดีที่สุด

กิตติกรรมประกาศ

ขอขอบคุณ รองศาสตราจารย์ อรุณ จีระวัฒน์กุล อาจารย์ประจำภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น ที่ให้คำแนะนำในการเขียนบทความนี้จนสำเร็จ

เอกสารอ้างอิง

1. Armitage P, Colton T. Multiple comparisons. Encyclopedia of Biostatistics 1998; 1:2736-45.
2. Beal GK, Khamis HJ. A problem in statistical analysis: simultaneous Inference. The Condor 1991; 93:1023-5.
3. Bewick V, Cheek L, Ball J. Statistics review9: one-way analysis of variance. Critical Care 2004; 8(2):130-6.
4. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995; 310:170.
5. Wassertheil-Smoller S. Biostatistics and epidemiology: a primer for health professionals. 2nd ed: New York: Springer-Verlag; 1995.
6. Zar JH. Biostatistics analysis. 3rd ed. Prentice-Hall; 1996.

Abstract Critical Points in Multiple Comparisons

Phorjai Pattanitham

Journal of Health Science 2008; 17:SV1225-31.

Analysis of variance (ANOVA) is a statistical method for comparing the population means of more than two groups. When a statistically significant difference across those population means has been found, a multiple comparison method can be used to look for specific differences between pairs of groups. This article presents the principle of a single-step procedure for multiple comparisons of the Bonferroni, Scheffe and Turkey method. The Bonferroni method is the most effective for comparing some pairwise comparisons while the most acceptable for testing all pairwise comparisons is Turkey method. Considering only the p-values obtained from the multiple comparisons is inadequate for interpretation of results. The confidence intervals for each population mean being compared and the confidence intervals for a difference between population means of the pairwises should be used to summarize the findings together with the scientifically meaningful in order to make a practical information for decisions to put the findings into practice.

Key words: critical points, bivariate analysis