

ปกิณกะวิจัย: ตัวเลขไร้สาระหรือตัวเลขไร้ปัญญา
และผลที่ตามมาเมื่อนำมาใช้
Simpson's Paradox and Its Consequences When
Applying It

พิชิต พัทธกะเทพสมบัติ¹

Pichit Pitaktepsombat

“การเริ่มต้นที่ผิดพลาด ถ้ายังดำเนินการขั้นตอนต่อไป ความผิดพลาด
ย่อมตามมาอย่างหลีกเลี่ยงไม่ได้”

คำสำคัญ: *Simpson's Paradox, Level of Measurement, Composite
Index, Mean*

1. คำนำ

คนโดยทั่วไปและโดยเฉพาะนักศึกษา นักวิชาการหรือนักวิจัยต้องเคยเห็นเคย
ใช้ข้อมูลหรือสถิติต่าง ๆ ซึ่งก็คงจะพอรู้พอเข้าใจได้ แต่ในบางครั้งอาจจะได้พบกับ
ตัวเลข สถิติต่าง ๆ ที่ทำให้งุนงง ไม่เข้าใจว่าหมายความว่าอะไร คำนวณมาทำอะไร
เช่น เมื่อมีการเสนอค่าเฉลี่ย (mean) ของเพศ เท่ากับ 0.35, 0.57 หรือ 0.85 เป็นต้น
สำหรับตัวแปรตัวเดียว SPSS สามารถคำนวณสถิติต่าง ๆ (เช่น Mean, Median,
Mode, Std. deviation, Variance ฯลฯ) ได้เสมอไม่ว่าตัวแปรจะอยู่ระดับกลุ่ม
(nominal level) อันดับ (ordinal level) ช่วง (interval level) หรืออัตราส่วน (ratio

¹ ศาสตราจารย์ คณะรัฐประศาสนศาสตร์ สถาบันบัณฑิตพัฒนบริหารศาสตร์
สถานที่ติดต่อ: เลขที่ 118 ถ.เสรีไทย คลองจั่น บางกะปิ กทม. 10240

level) เพราะรายการหรือค่าของคำถามหรือตัวแปรที่เป็น input ป้อนเข้าในคอมพิวเตอร์ไม่ใช่ตัวอักษร เช่น ชาย หญิง หรือเห็นด้วย ไม่เห็นด้วย หรือต่ำ ปานกลาง สูง ฯลฯ แต่เป็นตัวเลข 0 ถึง 9 ซึ่งตัวเลขเหล่านี้บางครั้งไม่มีความหมายในทางคณิตศาสตร์หรือไม่ใช้ตัวเลขที่แท้จริงแต่อย่างใด เป็นเพียงสัญลักษณ์หรือรหัสเท่านั้น บางครั้งก็เป็นตัวเลขจริงที่มีความหมายทางคณิตศาสตร์

เมื่อใดก็ตามที่นักวิจัยเข้าใจคลาดเคลื่อนว่าตัวเลขนั้นเป็นตัวเลขจริง ๆ ทั้ง ๆ ที่เป็นแค่รหัส หรือในกรณีที่ตัวเลขหรือรหัสนั้นเป็นตัวเลขที่แท้จริง แล้วเอาตัวเลขเหล่านี้มาหาค่าเฉลี่ย ก็จะพบกับตัวเลขหรือค่าเฉลี่ยที่ยากจะเข้าใจว่าหมายความว่าอะไร คำนวณมาทำอะไร ดังตัวอย่างค่าเฉลี่ยของเพศที่ได้ยกตัวอย่างมาแล้ว เมื่อนักวิจัยหรือผู้อ่านไม่เข้าใจก็เริ่มใช้ความรู้สึกนึกคิดส่วนตัว (subjective) เพื่อที่พยายามจะเข้าใจตัวเลขหรือค่าเฉลี่ยนั้นโดยไม่ได้อาศัยศาสตร์เข้ามาช่วย นักวิจัยจะต้องหลีกเลี่ยงวิธีปฏิบัติเช่นนี้ ไม่เช่นนั้นก็จะกลายเป็นว่านักวิจัยใช้ความรู้สึกนึกคิดส่วนตัวไปอธิบายปรากฏการณ์ต่าง ๆ ตามใจชอบ

มีนักวิจัยเสนอค่าเฉลี่ยประเภทนี้ เพราะทำตามคนอื่น ๆ ที่เคยทำมา ขาดการพินิจพิเคราะห์ จึงนำเสนอสถิติที่เรียกว่า Simpson's Paradox ซึ่งผู้เขียนจะขอแปลว่า "ตัวเลขไร้สาระหรือไร้ปัญญา"

2. วัตถุประสงค์

ปณิธานวิจัยนี้มีวัตถุประสงค์ที่สำคัญดังต่อไปนี้

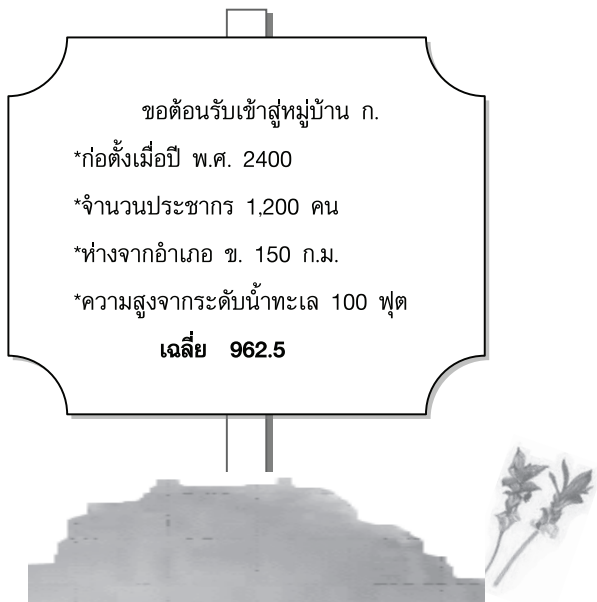
1. เพื่อเผยแพร่สถิติที่เรียกว่า Simpson's Paradox ให้เป็นที่ทราบและเข้าใจอย่างทั่วถึงมากยิ่งขึ้น
2. ชี้ให้เห็นผลที่จะตามมา เมื่อมีการนำสถิติประเภทนี้ไปใช้

3. Simpson's Paradox คืออะไร และผลที่ตามมาเมื่อนำมาใช้

ผู้เขียนเชื่อว่ามีคนจำนวนไม่น้อยที่ไม่เคยเห็นหรือไม่เคยได้ยิน หรือ อาจจะเคยได้ยินมาบ้าง ผู้เขียนเองก็เพิ่งจะได้พบคำนี้เมื่อไม่นานมานี้ จากหนังสือ Intro

Stats ของ De Veaux et al. เป็นชื่อที่ตั้งขึ้นเพื่อเป็นการให้เกียรติกับนักสถิติที่สร้างสถิติตัวนี้ในทศวรรษ 1960

De Veaux et al. ได้นำเสนอสถิตินี้โดยกล่าวว่า “จงอย่าใช้ค่าเฉลี่ยที่ไม่เหมาะสมหรือโง่ๆ (Don’t use unfair or silly averages)” เพราะการใช้หรือนำเสนอค่าเฉลี่ยนี้สามารถพาให้หลงผิดได้ เพราะฉะนั้นจึงต้องระมัดระวังเมื่อหาค่าเฉลี่ยดังตัวอย่างป้ายเข้าหมู่บ้าน ก.



อีกตัวอย่างหนึ่งที่จะขอนำเสนอคือ การบินตรงต่อเวลาของนักบิน ดังตาราง

นักบิน	เวลากลางวัน	เวลากลางคืน	รวม
ก.	90 ครั้ง จาก 100 ครั้ง	10 ครั้ง จาก 20 ครั้ง	100 ครั้งจาก 120 ครั้ง
	90%	50%	83%
ข.	19 ครั้ง จาก 20 ครั้ง	75 ครั้ง จาก 100 ครั้ง	94 ครั้งจาก 120 ครั้ง
	95 %	75%	78%

จากสถิติข้างบน นักบิน ข. สามารถบินได้ตรงต่อเวลาทั้งกลางวันและกลางคืน ดีกว่านักบิน ก. แต่สถิติรวมปรากฏว่านักบิน ก. มีสถิติดีกว่า เป็นไปได้อย่างไร ? ที่เป็นเช่นนี้เพราะข้อมูลหรือจำนวนครั้งของนักบิน ข. เป็นจำนวนมาก เป็นการบินในเวลากลางคืน ซึ่งยากกว่าการบินในเวลากลางวัน ส่วนนักบิน ก. ได้ประโยชน์จากการบินจำนวนมากเป็นเวลากลางวัน ซึ่งง่ายต่อการบิน จากรูปแบบของการบิน การนำมาหาค่าเฉลี่ยจึงเป็นการพาลหลงผิดได้ (misleading)

อีกกรณีตัวอย่างจริง ซึ่งเป็นที่รู้จักกันดีที่แสดงให้เห็น Simpson's Paradox เมื่อมีการสอบสวนอัตราการรับนักศึกษาชายต่อหญิงของ University of California at Berkeley ซึ่งนำเสนอในวารสาร Science ที่รายงานว่า ผู้สมัครชายได้รับเลือกร้อยละ 45.0 ในขณะที่ผู้สมัครหญิงได้รับเลือกเพียงร้อยละ 30.0 ซึ่งทำให้ดูเหมือนว่าการเลือกปฏิบัติ แต่อย่างไรก็ตามเมื่อแยกพิจารณาตามคณะที่สมัคร ปรากฏว่าแต่ละคณะ หญิงได้รับการคัดเลือกในสัดส่วนที่ใกล้เคียงกับชาย ในบางคณะหญิงได้รับการคัดเลือกมากกว่าชายเสียอีก ทำไมจึงเป็นเช่นนั้น ข้อเท็จจริงก็คือ หญิงเป็นจำนวนมากสมัครคณะที่มีอัตราการรับต่ำ เช่น คณะนิติศาสตร์ แพทยศาสตร์ (มีอัตราการรับต่ำกว่าร้อยละ 10) ชายมีแนวโน้มที่จะสมัครคณะวิศวกรรมศาสตร์ และ วิทยาศาสตร์ ซึ่งมีอัตราการรับสูงกว่าร้อยละ 50.0 เมื่อหาค่าเฉลี่ยรวมทุกคณะจึงปรากฏว่าหญิงได้รับการคัดเลือกต่ำกว่าชาย ค่าเฉลี่ยนี้จึงไร้สาระ ทำให้เข้าใจผิดในสาระสำคัญ

จาก 3 ตัวอย่างที่ได้นำเสนอมาแล้ว เป็นการหาค่าเฉลี่ยของตัวแปรระดับอัตราส่วนหลาย ๆ ตัวที่มีลักษณะต่างกัน นั่นคือ ปีที่ก่อตั้ง (พ.ศ. 2400) จำนวนประชากร (1,200 คน) ระยะห่างจากอำเภอ (150 ก.ม.) ความสูงจากระดับน้ำทะเล (100 ฟุต) ซึ่งเป็นตัวแปรระดับอัตราส่วน สามารถหาค่าเฉลี่ยได้ แต่ค่าเฉลี่ยนี้ไม่มีความหมายอะไร

นอกจาก 3 ตัวอย่างของ De Veaux et al. ผู้เขียนใคร่ขอเสนอกรณีอื่น ๆ ที่น่าจะเรียกได้ว่าเป็นตัวอย่างของ Simpson's Paradox นั่นคือ การหาและนำเสนอค่าเฉลี่ยของตัวแปรเดียว (Univariate analysis) ที่อยู่ระดับกลุ่มและระดับอันดับ ซึ่งเป็นกรณีที่พบเห็นเสมอ ๆ จนกลายเป็นการปฏิบัติที่คิดว่าถูกต้องทั้ง ๆ ที่ไม่ถูกต้อง

เพราะตัวแปรตัวเดียวหรือคำถามเดียวที่อยู่ระดับกลุ่มและระดับอันดับนั้น สถิติที่คำนวณแล้วนำมาใช้ได้คือ ร้อยละ ฐานนิยม และมัธยฐาน ตามลำดับ เท่านั้น ไม่สามารถคำนวณค่าเฉลี่ยได้ แต่อย่างไรก็ตามเมื่อนำเอาตัวแปรระดับอันดับตั้งแต่ 2 ตัวขึ้นไปมารวมกัน เพื่อสร้าง Composite index (โดยการ compute จากหลาย ๆ ตัวแปรหรือคำถามให้กลายเป็นแนวคิด) ให้นุโลมว่าตัวแปรที่สร้างขึ้นใหม่นั้นเป็นตัวแปรระดับอัตราส่วนจึงสามารถหาค่าเฉลี่ยได้

ทำไมจึงเกิดความเข้าใจผิดหรือคลาดเคลื่อนว่าคำถามหรือตัวแปรแต่ละข้ออยู่ที่ระดับช่วงหรืออัตราส่วนเช่นนี้ ผู้เขียนคิดว่าน่าจะมีคำอธิบาย นั่นคือ เกิดจากการออกแบบแบบสอบถามหรือแบบสัมภาษณ์ซึ่งจะได้แสดงให้เห็นเป็นตัวอย่าง

ตัวอย่าง (ที่ผิด ๆ)

แบบสอบถามเกี่ยวกับคุณภาพการให้บริการ

คำชี้แจง กรุณาใส่เครื่องหมาย ✓ ลงในช่องว่างทางขวามือเพียงช่องเดียว ตามระดับความคิดเห็นของท่าน โดยแบ่งระดับความเห็นออกเป็น 5 ระดับ ดังนี้

5 = มากที่สุด, 4 = มาก, 3 = ปานกลาง, 2 = น้อย, 1 = น้อยที่สุด

ข้อ	คำถาม	5 มากที่สุด	4 มาก	3 ปานกลาง	2 น้อย	1 น้อยที่สุด
1.	ผู้ให้บริการแสดงความพร้อมในการให้บริการต่อท่านเสมอ					
2.	ท่านได้รับการบริการที่เกิดความคาดหวังเสมอ					
3.	ผู้ให้บริการมักแสดงพฤติกรรมที่ไม่สนใจในการให้บริการกับท่าน					
⋮						

หมายเหตุ : จากตัวอย่าง ไม่มีความจำเป็นใดที่จะต้องใส่ตัวเลข เพราะมีคำตอบอยู่แล้ว การใส่ตัวเลขเมื่อผู้วิจัยเห็นทำให้เข้าใจผิดว่าเป็นตัวเลขที่แท้จริง

รูปแบบที่พบเห็นเป็นประจำและบ่อยกว่าที่นำเสนอข้างบนคือ คำตอบ มีเพียงตัวเลข 5 4 3 2 1 เท่านั้น ไม่มีคำตอบ “มากที่สุด, ...” เหตุที่เป็นเช่นนี้น่าจะเกิดจากปัญหาเนื้อที่ที่จะพิมพ์คำตอบ “มากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด” ไม่เพียงพอ เพราะทางด้านแนวนอนต้องพิมพ์คำถามแต่ละข้อลงไป ผู้พิมพ์มักจะพิมพ์คำถามก่อนจึงไม่เหลือพื้นที่สำหรับคำตอบในแนวตั้ง ที่จะให้พิมพ์คำตอบที่เป็น

ตัวอักษรได้ จึงต้องใช้ 5, 4, 3, 2, 1 แทน การใช้ตัวเลขแบบนี้ก่อให้เกิดปัญหาแก่ผู้ตอบ นั่นคือ เมื่อผู้ตอบต้องการตอบว่า “เห็นด้วยอย่างยิ่ง” หรือ คำตอบอื่นๆ ไม่ใช่คำตอบ 5, 4, ... เพราะฉะนั้นเวลาผู้ตอบตอบคำถามเขาก็ต้องไปดูว่า 5, 4, 3, 2, 1 คืออะไร จึงเป็นการสร้างภาระให้ผู้ตอบ

ผู้เขียนก็เคยปฏิบัติมาเช่นนี้เพราะข้อจำกัดในเรื่องเนื้อที่ แต่ในการวิจัย เนื้อหาเป็นเรื่องสำคัญกว่ามาก ผู้เขียนจึงได้ขจัดข้อจำกัด (น่าจะเรียกได้ว่า “อับปัญญา”) นี้ โดยใส่คำตอบ “มากที่สุด” “มาก” “ปานกลาง” “น้อย” และ “น้อยที่สุด” หรือคำตอบอื่น ๆ ที่เป็นเรื่องเนื้อหา แทนที่จะใส่ตัวเลข ดังตัวอย่าง จึงไม่ก่อให้เกิดความเข้าใจผิด

คำกล่าว	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด	ไม่ทราบ / ไม่ตอบ
1. ผู้ให้บริการแสดงความพร้อมในการให้บริการต่อท่านเสมอ						
2. ท่านได้รับการบริการที่เกินความคาดหวังเสมอ						
3. ผู้ให้บริการมักแสดงพฤติกรรมที่ไม่สนใจในการให้บริการกับท่าน						
⋮						

หมายเหตุ : เมื่อไม่มีตัวเลข ก็จะไม่นำไปสู่ความเข้าใจผิด

4. ผลที่ตามมา

ไม่ว่าจะเป็นการคำนวณค่าเฉลี่ยจากตัวแปรระดับช่วงหรืออัตราส่วน ซึ่งมีลักษณะต่างกัน เช่น ระยะทาง น้ำหนัก ฯลฯ ซึ่งสามารถทำได้ หรือจากตัวแปรระดับกลุ่มและระดับอันดับ ซึ่งไม่สามารถทำได้ แต่ผู้วิจัยมักจะทำกันอันเป็นผลมาจากความเข้าใจผิด ๆ หรือคลาดเคลื่อน ผลที่ตามมาสำคัญคือ

1. ในกรณีของตัวแปรระดับช่วงหรืออัตราส่วนที่มีลักษณะต่างกัน เช่น อายุ รายได้ เป็นต้น ค่าเฉลี่ยที่ได้ไม่มีความหมายใด ๆ แถมยังจะก่อให้เกิดความงุนงงอีก

2. ในกรณีของตัวแปรระดับกลุ่มหรืออันดับ สถิติที่ SPSS คำนวณมาให้สถิติที่ใช้ได้และเหมาะสมคือ ร้อยละ ฐานนิยม (เมื่อเป็นตัวแปรระดับกลุ่ม) ร้อยละ และ

มัธยฐาน (เมื่อตัวแปรอยู่ระดับอันดับ) **ไม่ใช่ค่าเฉลี่ย** สำหรับตัวแปรตัวเดียวไม่ว่าจะอยู่ระดับใดก็ตาม SPSS จะให้สถิติทั้งหมด 11 ตัวเหมือนกัน จึงเป็นหน้าที่ของผู้วิจัยจะต้องเลือกใช้ให้ถูกต้องและเหมาะสม ไม่ใช่ใช้ได้ทุกตัวที่ SPSS คำนวณ และพิมพ์ออกมาให้

เมื่อเข้าใจผิดหรือคลาดเคลื่อน จึงเลือกใช้ค่าเฉลี่ยซึ่งเป็นการเริ่มต้นที่ผิด การกระทำขั้นตอนต่อไปย่อมผิดต่อไปและน่าจะมากขึ้นด้วย ดังจะแสดงให้เห็นจากตัวอย่างรายงานการวิจัยที่ได้นำเสนอมาก่อน

ส่วนที่ 1 ข้อมูลในการวัดความพึงพอใจด้านความคิดเห็นของลูกค้า ในการใช้บริการล้างทำความสะอาดรถยนต์ ซึ่งเป็น 5 ระดับดังนี้		
ระดับที่ 1 พอใจมากที่สุด	คะแนน* 5 คะแนน	
ระดับที่ 2 พอใจมาก	คะแนน* 4 คะแนน	
ระดับที่ 3 พอใจปานกลาง	คะแนน* 3 คะแนน	
ระดับที่ 4 พอใจน้อย	คะแนน* 2 คะแนน	
ระดับที่ 5 พอใจน้อยที่สุด	คะแนน* 1 คะแนน	
ความหมายของค่าเฉลี่ยทั้ง 5 ระดับ ใช้เกณฑ์ประเมินดังนี้		
ค่าเฉลี่ย 4.51 - 5.00	ถือว่า มีความพึงพอใจมากที่สุด	} เกณฑ์ที่สร้าง มาภายหลัง
ค่าเฉลี่ย 3.51 - 4.50	ถือว่า มีความพึงพอใจมาก	
ค่าเฉลี่ย 2.51 - 3.50	ถือว่า มีความพึงพอใจปานกลาง	
ค่าเฉลี่ย 1.51 - 2.50	ถือว่า มีความพึงพอใจน้อย	
ค่าเฉลี่ย 1.00 - 1.50	ถือว่า มีความพึงพอใจน้อยที่สุด	

ข้อสังเกตจากตัวอย่างท่อนแรก คำตอบในแบบสอบถามนั้นคือ มากที่สุด ... ที่ผู้วิจัยเรียกเป็นคะแนนซึ่งก็ไม่ใช่คะแนนในความหมายทางคณิตศาสตร์ เพราะฉะนั้นการเรียกว่าเป็นคะแนนจึงเป็นการหลงผิดในเนื้อหา แต่เมื่อมีความจำเป็นที่จะต้องรวม

* คำว่า**คะแนน** (ซึ่งมี 1-5) ดังตัวอย่างด้านบน เป็นการใช้ภาษาหรือคำที่ผิดหรือคลาดเคลื่อน เพราะจริง ๆ แล้ว ไม่ใช่คะแนน เป็นเพียงรหัส หรือสัญลักษณ์ที่ใช้แทนคำตอบ ‘พอใจน้อยที่สุด’ซึ่งถ้าทำตามแบบอย่างของผู้เขียน ก็จะไม่มีความจำเป็นต้องมีตัวเลขในแบบสอบถาม แต่เมื่อป้อนข้อมูลลงใน SPSS เป็นการสะดวกมากกว่า ถ้าจะใช้ตัวเลข 1 แทนน้อยที่สุด...5 แทนมากที่สุด และสุดท้ายก็ไม่จำเป็นต้องมีข้อความที่ทำให้หลงผิดในกรอบด้านบนเลย

คำถามหรือตัวแปรเหล่านี้เพื่อที่จะกลับไปสู่แนวคิด ซึ่งในที่นี้คือความพึงพอใจ โดยการให้คะแนนตามท่อนแรกก็ทำได้ แต่พอเปลี่ยนมาเป็นเกณฑ์ในการประเมิน (ผู้เขียนก็ใคร่ถามว่าเอาไปประเมินอะไร) เกิดความไม่สอดคล้องกัน นั่นคือ ในท่อนบน คะแนน 1 = พอใจน้อยที่สุด แต่ในท่อนล่างเปลี่ยนเป็นค่าเฉลี่ย 1.00 – 1.50 ผู้เขียนเชื่อว่าความไม่สอดคล้องนี้เกิดจากความไม่ระมัดระวังในการใช้คำมากกว่า จึงดูเหมือนว่าต่างกันที่จริงก็เรื่องเดียวกันนั่นเองเพียงแค่คะแนนต่างกัน จะเห็นได้ว่าผู้วิจัยขาดความตระหนักในสิ่งที่ตนทำ ทั้ง ๆ ที่เป็นเรื่องเดียวกัน ห่างกันไม่กี่บรรทัด

จากประสบการณ์ของผู้เขียนได้พบว่า นักวิจัยได้นำเสนอเกณฑ์ที่แตกต่างกัน ผู้เขียนก็ไม่แน่ใจว่าได้เกณฑ์อย่างไร ใช้เกณฑ์ของใคร (บางครั้งได้รับคำตอบว่า “ของพวกผมเอง”) จึงได้พยายามค้นหาที่มาของเกณฑ์เช่นนี้จากนักวิชาการต่างประเทศและของไทย (ที่เชื่อถือได้) ก็ยังไม่พบ ที่พบในรายงานฉบับหนึ่งได้มีการอ้างอิงเกณฑ์จากหนังสือ Best, J.W. 1978. *Research in Education*. 3rd. New York: Prentice Hall. ก็พบว่าเกณฑ์ของ Best นั้นใช้กับค่าของ Correlation Coefficient ซึ่งเป็นเรื่องของ 2 ตัวแปร ไม่ใช่ตัวแปรตัวเดียว จึงเป็นการอ้างอิงผิด ๆ นอกเสียจากได้เปลี่ยนค่านั้นให้เท่ากับ 0 ถึง 1.0

5. สรุป

Simpson's Paradox คือ ค่าเฉลี่ยที่เกิดจากการนำคำถามหลาย ๆ คำถามหรือตัวแปรหลายตัวแปรที่อยู่ระดับช่วงหรืออัตราส่วนที่มีลักษณะแตกต่างกันมาคำนวณเป็นค่าเฉลี่ย หรือค่าเฉลี่ยที่เกิดจากการนำเอาคำถามเดียวหรือตัวแปรเดียวที่อยู่ระดับกลุ่มหรืออันดับมาหาค่าเฉลี่ย เท่านั้นยังไม่พอนักวิจัยจำนวนไม่น้อยยังพยายามหาความหมายจากความผิดพลาดนี้ สถิติที่ได้ไม่มีความหมายอะไร ซึ่งเรียกว่า “Simpson's Paradox”.

เอกสารอ้างอิง

พิชิต พิทักษ์เทพสมบัติ (บรรณาธิการ). (2551). จงอย่าใช้ตัวแปรระดับอันดับในฐานะตัวแปรระดับช่วงหรืออัตราส่วน ในงทำและจงอย่าทำในการวิจัยสังคม. กรุงเทพฯ: เสมาธรรม.

De Veaux, R.D., et al. (2006). *Intro Stats*. New York: Pearson/Addison Wesley.