

# การใช้สถิติสัมประสิทธิ์สหสัมพันธ์อย่างไม่เหมาะสม

อรุณ จิรวัดนกุล

งานวิจัยทางด้านพฤติกรรมสุขภาพ มีการใช้สถิติสัมประสิทธิ์สหสัมพันธ์วิเคราะห์หาความสัมพันธ์ระหว่างตัวแปร จากบทความวิจัยที่ตีพิมพ์ยังพบว่ามีการใช้สัมประสิทธิ์สหสัมพันธ์อย่างไม่เหมาะสมหลายกรณี ที่มีผลต่อความถูกต้องของข้อสรุป

สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรคู่  $x$  และ  $y$  คำนวณจากสัดส่วนความแปรปรวนร่วมของ  $x$  และ  $y$  หารด้วยผลคูณของส่วนเบี่ยงเบนมาตรฐานของ  $x$  และส่วนเบี่ยงเบนมาตรฐาน  $y$

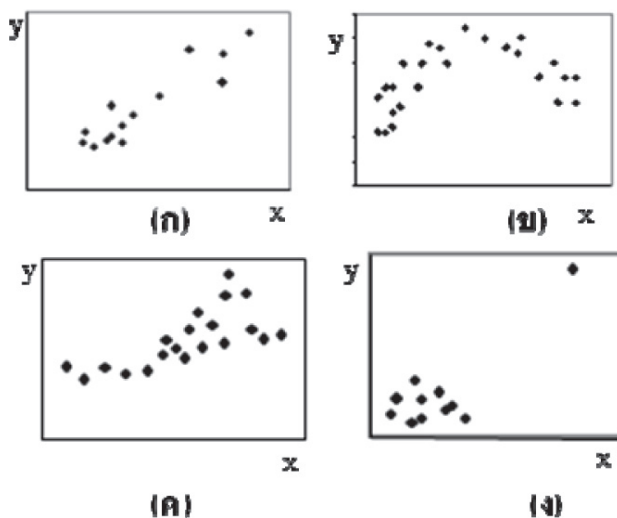
$$\text{Correlation}(X, Y) = \frac{\text{Covariance}(X, Y)}{\sqrt{\text{Variance}(X) \times \text{Variance}(Y)}}$$

ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณจากสูตรดังกล่าวใช้แสดงความสัมพันธ์เชิงเส้นตรงระหว่างสองตัวแปรโดยตัวแปร  $x$  และ  $y$  ต้องเป็นข้อมูลต่อเนื่องมาจากหน่วยศึกษาเดียวกันและมีการแจกแจงปกติ (bivariate normal distribution)<sup>(1)</sup> ค่าสัมประสิทธิ์สหสัมพันธ์ใช้วัดว่าสองตัวแปรมีความสัมพันธ์เชิงเส้นตรงในระดับใด ความไม่เหมาะสมของการใช้สัมประสิทธิ์สหสัมพันธ์ ( $r$ ) ที่พบมากที่สุดคือการนำเสนอค่า  $r$  และผลการทดสอบว่ามีนัยสำคัญโดยไม่แสดงแผนภาพกระจาย (scatter plot)<sup>(2)</sup> หรือไม่ระบุการตรวจสอบว่าลักษณะความสัมพันธ์ เป็นเชิงเส้นตรงหรือไม่ โปรแกรมวิเคราะห์ทางสถิติคำนวณค่า  $r$  โดยไม่ได้ดูว่าตัวแปรทั้งสองมีลักษณะความสัมพันธ์เป็นเชิงเส้นตรงหรือไม่ ดังนั้นถ้านักวิจัยดู

เฉพาะค่า  $r$  และ  $p$ -value ที่คำนวณได้แล้วนำมาสรุปผลงานวิจัยโดยไม่พิจารณาแผนภาพกระจายว่าตัวแปรทั้งสองมีลักษณะความสัมพันธ์เป็นแบบใดอาจทำให้สรุปความสัมพันธ์ผิด

จากรูปที่ 1(ก) แผนภาพกระจายแสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงใช้  $r$  สรุปความสัมพันธ์ได้ ส่วนรูปที่ 1(ข) ตัวแปรทั้งสองไม่ได้สัมพันธ์เชิงเส้นตรงถึงแม้ได้ค่า  $r = 0.4$   $p$ -value  $< 0.001$  จะใช้  $r$  สรุปความสัมพันธ์ไม่ได้

ความสัมพันธ์ของข้อมูลที่มีลักษณะต่างกัน เป็นช่วง ๆ ดังในรูปที่ 1(ค) ในช่วงที่ ตัวแปร  $x$  มีค่าน้อยจะไม่เหมือนกับช่วง  $x$  มีค่ามาก (heteroscedasticity) ความสัมพันธ์ในลักษณะนี้ไม่เหมาะที่จะใช้  $r$  แสดง



รูปที่ 1 แผนภาพกระจายแสดงความสัมพันธ์สองตัวแปร

ความสัมพันธ์ สถิติ  $r$  เหมาะสมกับข้อมูลในทุกช่วงมีลักษณะความสัมพันธ์เหมือนกัน (homoscedasticity) เช่นรูปที่ 1 (ก)

รูปที่ 1 (ง) ข้อมูลที่ศึกษาบางค่ามีค่าห่างผิดปกติจากกลุ่ม (outliers) รวมอยู่ด้วยจะมีผล กระทบต่อค่า  $r$  ที่คำนวณได้ ทั้งที่จะทำให้ค่า  $r$  มีค่ามากขึ้น หรือ  $r$  มีค่าเล็กลง

จะเห็นได้ว่าลักษณะความสัมพันธ์มีผลต่อการแปลผลค่า  $r$  ดังนั้นนักวิจัยต้องสั่งให้โปรแกรมวิเคราะห์สร้างแผนภาพกระจายเพื่อใช้ประกอบการพิจารณาค่า  $r$  และ  $p$ -value จะทำให้ข้อสรุปที่ได้ถูกต้อง ในการนำเสนอผลการศึกษาที่ใช้  $r$  แสดงแสดงความสัมพันธ์นักวิจัยควรนำเสนอแผนภาพกระจายประกอบด้วย

อีกกรณีหนึ่งของการใช้ไม่เหมาะสมคือการใช้สถิติ  $r$  แสดงค่าความสอดคล้อง (agreement) ของการวัดผลสองวิธี เช่น การศึกษาความสอดคล้องของผลการวัดความดันโลหิตสูงด้วยเครื่องอัตโนมัติกับการวัดโดยพยาบาล ความสอดคล้องของการวัดเป็นการดูว่าผลการวัดของคน ๆ เดียวกัน ที่วัดด้วยเครื่องอัตโนมัติ (x) สอดคล้อง (เท่ากับหรือใกล้เคียง) กับผลการวัดโดยพยาบาล (y) หรือไม่ สถิติที่ใช้วัดความสอดคล้องคือ interclass correlation หรือ concordance correlation coefficient

การศึกษาความสัมพันธ์ของตัวแปรที่ได้จากการวัดซ้ำ (repeated measure) ซึ่งมีรูปแบบความสัมพันธ์แบบมีเงื่อนไข (condition) ดังนั้นจึงไม่เหมาะที่จะใช้  $r$  ในการวิเคราะห์ความสัมพันธ์ ต้องใช้ concordance correlation coefficient หรือ mixed effects model วิเคราะห์ความสัมพันธ์ที่ได้จากการวัดซ้ำ

การอธิบายผลความสัมพันธ์ที่วิเคราะห์ด้วยสถิติ  $r$

นอกจากระดับความสัมพันธ์แล้วนักวิจัยควรอธิบายว่าความสัมพันธ์ที่พบนี้เป็นเชิงสาเหตุและผลหรือเป็นความสัมพันธ์ทางอ้อมที่ตัวแปรทั้งคู่ไปมีความสัมพันธ์กับตัวแปรอีกตัวหนึ่ง เช่น ในการศึกษาคุณภาพชีวิตของคนงาน พบว่าสมรรถนะการทำงานมีความสัมพันธ์ระดับสูงกับคุณภาพชีวิต นักวิจัยจึงเสนอให้พัฒนาสมรรถนะการทำงาน เพื่อเพิ่มคุณภาพชีวิตให้ดีขึ้น

ในความเป็นจริงสมรรถนะการทำงานมีความสัมพันธ์กับเงินเดือน และคุณภาพชีวิตก็มีความสัมพันธ์กับเงินเดือนจึงมีผลทำให้สมรรถนะการทำงานมีความสัมพันธ์กับคุณภาพชีวิต แต่การเสนอให้พัฒนาสมรรถนะการทำงานให้ดีขึ้น โดยไม่เพิ่มเงินเดือนตามสมรรถนะการทำงานที่เพิ่มขึ้นคุณภาพชีวิตจะไม่ดีขึ้นตามสมรรถนะการทำงาน

**สรุป** ในการนำเสนอผลการวิเคราะห์ด้วยสถิติสัมพันธ์สหสัมพันธ์ต้องนำเสนอคู่กับแผนภาพกระจายเพื่อให้แปลผลได้ถูกต้อง การวัดความสอดคล้องและการวัดซ้ำไม่เหมาะที่จะใช้สถิติสัมพันธ์สหสัมพันธ์ในการวิเคราะห์ การแปลผลสถิติสัมพันธ์สหสัมพันธ์ต้องพิจารณาว่าเป็นความสัมพันธ์ในลักษณะใด จะช่วยให้การแปลผลและให้ข้อเสนอแนะได้ถูกต้อง

#### เอกสารอ้างอิง

1. อรุณ จิรวัดน์กุล. สถิติทางวิทยาศาสตร์สุขภาพเพื่อการวิจัยที่ใช้ในงานวิจัยทางวิทยาศาสตร์สุขภาพ. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร: วิทยพัฒน์ ; 2552.
2. Porter AMW. Misuse of correlation and regression in three medical journals. R Soc Med 1999;92:123-8.