

การแปลงข้อมูลให้ตัวแปรมีการแจกแจงปกติ

อรุณ จิรวัดนกุล วท.บ. (อาชีวอนามัย), วท.ม. (ชีวสถิติ), M.Sc. (Clinical Epidemiology)

การใช้สถิติวิเคราะห์ทดสอบสมมติฐานเพื่อสรุปผลการเปรียบเทียบจากตัวอย่างไปเป็นข้อสรุปของประชากรโดยใช้สถิติ parametric เช่น t-test ที่มีข้อตกลงเบื้องต้น (assumption) ว่าตัวแปรที่จะนำมาทดสอบต้องมีการแจกแจงปกติ ดังนั้น ก่อนที่จะทดสอบสมมติฐานตัวแปรใด นักวิจัยต้องดำเนินการตรวจสอบก่อนว่า ตัวแปรดังกล่าวมีการแจกแจงปกติหรือไม่ ถ้าพบว่าตัวแปรที่จะทดสอบสมมติฐานมีการแจกแจงไม่ใช่แบบปกติ นักวิจัยส่วนใหญ่จะใช้สถิติ nonparametric ที่มีอำนาจการทดสอบ (power of test) ต่ำมาทดสอบแทน การใช้สถิติที่มีอำนาจการทดสอบต่ำมาทดสอบ ทำให้ความสามารถระบุความต่างของการเปรียบเทียบได้น้อยลง อาจทำให้ได้ผลสรุปที่ผิดพลาด

วิธีที่ควรใช้ในการวิเคราะห์ คือ การแปลงข้อมูลตัวแปรดังกล่าวให้มีการแจกแจงปกติ แล้วนำข้อมูลแปลงไปทดสอบสมมติฐานด้วยสถิติ parametric ผลสรุปของการทดสอบที่ได้ สามารถใช้สรุปผลวิจัยเสมือนกับการวิเคราะห์ข้อมูลก่อนแปลงโดยมีอำนาจการทดสอบเท่าเดิม กรณีข้อมูลที่ไม่สามารถแปลงเป็นการแจกแจงปกติได้ หรือตัวอย่างขนาดเล็ก จึงจะใช้สถิติ nonparametric ทดสอบสมมติฐาน

การทดสอบสมมติฐานด้วยสถิติ parametric จะคำนวณค่าสถิติจากข้อมูลจริง ซึ่งข้อมูลจริงแต่ละตัวจะต่างกันมากหรือน้อยตามค่าของข้อมูล เช่น ข้อมูล ตัวที่ 1 ต่างจาก ข้อมูลตัวที่ 2 อยู่ 8 (20-12) (ดังแสดงในตาราง)

ข้อมูลตัวที่	1	2	3	4	5	6
ค่าข้อมูลจริง	12	20	32	54	88	75
ค่าอันดับ	1	2	3	4	6	5

ตัวที่ 3 ต่างจากตัวที่ 2 อยู่ 12 (32-20) ฯลฯ ส่วน nonparametric จะใช้อันดับที่ใช้ในการคำนวณค่าสถิติ เช่น ค่าอันดับตัวที่ 1 ต่างจากตัวที่ 2 อยู่ 1 และค่าอันดับตัวที่ 1 ต่างจากตัวที่ 3 อยู่ 2 ฯลฯ ในการวิเคราะห์แบบ nonparametric ความต่างระหว่างข้อมูลแต่ละตัว จะลดลงเหลือเป็นความต่างระหว่างอันดับ ทำให้อำนาจการระบุความต่างลดลงไปด้วย ดังนั้น เพื่อให้การทดสอบมีอำนาจการทดสอบเท่าเดิม จึงต้องแปลงข้อมูลให้มีการแจกแจงปกติแล้วทดสอบด้วยสถิติ parametric นอกจากนี้ในการหาขนาดความต่างของค่าเฉลี่ยระหว่างกลุ่ม ขนาดความต่างที่คำนวณได้จากข้อมูลแปลง สามารถแปลงกลับไปเป็นขนาดความต่างของข้อมูลจริง ความต่างที่แปลงกลับมาใช้นำไปใช้พิจารณาขนาดความต่างที่มีความสำคัญ (clinical meaningful) เพื่อใช้ช่วยตัดสินใจในการนำผลงานวิจัยไปใช้งาน ส่วน nonparametric ทำไม่ได้

การทดสอบสมมติฐานด้วยสถิติ parametric จะมีข้อกำหนดข้อตกลงเบื้องต้นว่าการแจกแจงของตัวแปรที่จะทดสอบต้องมีคุณสมบัติอย่างไร ถ้าคุณสมบัติของตัวแปรที่จะทดสอบ ไม่เป็นไปตามข้อกำหนดดังกล่าว ผลสรุปที่ได้จากการทดสอบจะคลาดเคลื่อน หรือไม่ถูกต้อง

ข้อตกลงเบื้องต้นของการทดสอบทางสถิติที่เกี่ยวข้องกับการแปลงข้อมูลได้แก่

1. Normality ตัวแปรต้องมีการแจกแจงแบบปกติ (t test)
2. Homogeneity of variance ข้อมูลของกลุ่มต่างๆ ที่จะเปรียบเทียบต้องมีความแปรปรวนเท่ากัน (ANOVA)
3. Linearity มีความสัมพันธ์เชิงเส้นตรง (linear regression)

ดังนั้นการตรวจสอบคุณสมบัติเบื้องต้น ก่อนการวิเคราะห์ทางสถิติจึงเป็นสิ่งจำเป็น

การแปลงข้อมูลคือวิธีการที่ใช้แปลงข้อมูลจากรูปแบบ (format) หนึ่งไปเป็นอีกรูปแบบหนึ่ง เช่น นักวิจัยต้องการแปลงอุณหภูมิที่วัดเป็นองศา Celsius ไปเป็นองศา Fahrenheit วิธีการที่ใช้คือสมการคณิตศาสตร์ $F = C \times 9/5 + 23$ จากสมการ เมื่อแทนค่า C เท่ากับ 37 ลงในสมการ ค่า F ที่คำนวณได้ 98.6 จะเป็นค่าแปลงเมื่อ C เท่ากับ 37 ค่าแปลงจากสมการนี้สามารถแปลงกลับไปเป็นข้อมูลเดิมได้ค่าเท่าเดิม การแปลงด้วยสมการนี้ใช้ได้ทั้ง linear และ non-linear เช่น ในการวิเคราะห์หาค่าเฉลี่ยของ antibodies ซึ่งมีการแจกแจงแบบ non-linear การหาค่าเฉลี่ยจะต้องแปลงค่า antibodies ที่วัดได้ของแต่ละคนเป็นค่า log นำค่า log ที่แปลงได้ไปหาค่าเฉลี่ย นำค่าเฉลี่ยที่คำนวณได้แปลงกลับ (anti-log) มาเป็นค่าเฉลี่ย antibodies ของตัวอย่างที่ศึกษา

การแปลงข้อมูลเป็นเพียงการเปลี่ยนหน่วยวัดเดิมเป็นหน่วยใหม่ (re-scaling) ระยะห่างของค่าข้อมูลแปลง

จะสัมพันธ์กับข้อมูลเดิม เมื่อแปลงกลับเป็นข้อมูลตามหน่วยวัดเดิมค่าความห่างของข้อมูลจะเท่าเดิม ดังนั้นการแปลงข้อมูลให้มีการแจกแจงปกติ ก่อนนำไปทดสอบสมมติฐานด้วยสถิติ parametric จะสามารถนำผลการทดสอบไปสรุปได้โดยตรง เสมือนกับการวิเคราะห์ข้อมูลก่อนแปลง

วิธีการแปลงข้อมูลให้มีการแจกแจงแบบปกติมีหลายวิธี เช่น logarithmic transformation square root transformation arcsine transformation ฯลฯ จะเลือกใช้วิธีใดขึ้นอยู่กับลักษณะการแจกแจงของข้อมูลก่อนแปลง

นักวิจัยควรกำหนดการแปลงข้อมูลลงในแผนการวิเคราะห์ด้วย เช่น “เปรียบเทียบผลต่างของความเจ็บปวดระหว่างกลุ่มด้วย independent t-test ก่อนการทดสอบสมมติฐาน จะวิเคราะห์ลักษณะการแจกแจง ถ้าพบว่าการแจกแจงไม่ใช่แบบปกติ จะแปลงข้อมูลให้มีการแจกแจงปกติก่อนการทดสอบ ถ้าไม่สามารถแปลงข้อมูลให้มีการแจกแจงแบบปกติได้ จะใช้ Mann Whitney U test ในการทดสอบ”

สรุป

การทดสอบสมมติฐานที่มีข้อตกลงเบื้องต้นว่าตัวแปรต้องมีการแจกแจงปกติ นักวิจัยควรตรวจสอบลักษณะการแจกแจงก่อนทำการทดสอบ ถ้าพบว่าตัวแปรที่จะทดสอบมีการแจกแจงที่ไม่ใช่แบบปกติ ควรแปลงข้อมูลก่อนใช้สถิติ parametric ทดสอบ ในกรณีตัวอย่างขนาดเล็ก หรือไม่สามารถแปลงให้มีการแจกแจงปกติได้ จึงจะใช้สถิติ nonparametric ในการทดสอบ