

ขนาดของความแตกต่างในการทดสอบเพื่อเปรียบเทียบค่าเฉลี่ย ของประชากรสองกลุ่ม

Effect Size in Statistical Test Comparing Two Population Means

สุรเมศวร์ ฮาซิม

ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต ปทุมธานี 12121

1. บทนำ

การศึกษาเพื่อเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของประชากรสองกลุ่มที่เป็นอิสระกัน หลักการทางสถิติที่ใช้ในการทดสอบคือ การเปรียบเทียบความแตกต่างของข้อมูลกลุ่มตัวอย่างทั้งสองกลุ่มว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติหรือไม่ ผลสรุปของการทดสอบจะแสดงถึงโอกาสที่ข้อมูลทั้งสองกลุ่มจะแตกต่างกันตามทีคาดหวังไว้ แต่ไม่ได้แสดงขนาดของความแตกต่างของข้อมูลทั้งสองกลุ่ม และไม่สามารถสรุปถึงประโยชน์ที่เกิดขึ้นจากความแตกต่างนี้ได้อย่างแท้จริง [1] นอกจากนี้ปัญหาที่สำคัญในการทดสอบการมีนัยสำคัญทางสถิติคือการเข้าใจที่ผิดพลาดจากผลสรุปที่ได้คือ 1) หากค่าพีที่ได้จากการทดสอบมีค่าน้อยแสดงว่าข้อมูลทั้งสองกลุ่มมีความแตกต่างกันมาก และ 2) เมื่อปฏิเสธสมมติฐานหลักแสดงว่าประชากรทั้งสองกลุ่มมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติเท่านั้น ไม่สามารถสรุปได้ว่าขนาดของความแตกต่างนั้นมีความสำคัญ [2]

การทดสอบความมีนัยสำคัญทางสถิติไม่ได้พิจารณาขนาดของความแตกต่างของข้อมูลทั้งสอง

กลุ่มเพียงอย่างเดียว แต่จะพิจารณาความคลาดเคลื่อนมาตรฐานของข้อมูลประกอบด้วย ทำให้ในการทดสอบบางครั้งแม้ว่าข้อมูลทั้งสองกลุ่มมีความแตกต่างกันมากแต่ก็ไม่มีนัยสำคัญทางสถิติ หรือในบางกรณีที่มีขนาดความแตกต่างของข้อมูลทั้งสองกลุ่มน้อย แต่กลับมีนัยสำคัญทางสถิติ ดังนั้นในการเปรียบเทียบความแตกต่างของประชากรสองกลุ่ม เมื่อผลการทดสอบปรากฏว่าประชากรทั้งสองกลุ่มไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ เราสามารถสรุปได้หรือไม่ว่าประชากรทั้งสองกลุ่มไม่มีความแตกต่างกันหรือมีความแตกต่างกันเพียงเล็กน้อยหรือเป็นเพราะสาเหตุอื่นๆ บทความนี้จะนำเสนอวิธีการศึกษาเปรียบเทียบความแตกต่างของค่าเฉลี่ยของประชากรสองกลุ่มที่เป็นอิสระกัน โดยพิจารณาขนาดความแตกต่างมาตรฐานของข้อมูลทั้งสองกลุ่มตามแนวคิดของ Cohen [อ้างอิงใน 3] และการทดสอบความแตกต่างอย่างมีนัยสำคัญทางสถิติ เพื่อใช้ประกอบร่วมกันในการศึกษาหาข้อสรุปเพื่อเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของประชากรสองกลุ่ม

2. การทดสอบความแตกต่างค่าเฉลี่ยของประชากรสองกลุ่มอย่างมีนัยสำคัญทางสถิติ

ในการทดสอบสมมติฐานเพื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยของประชากรสองกลุ่มที่เป็นอิสระกันด้วยกระบวนการทางสถิติ จะมีข้อตกลงเบื้องต้นคือ ประชากรทั้งสองกลุ่มต้องมีการแจกแจงแบบปกติ และทราบค่าความแปรปรวนของประชากรทั้งสองกลุ่ม โดยทั่วไปแล้วจะมีขั้นตอนในการทดสอบดังนี้

1. กำหนดสมมติฐานของการทดสอบ

$$H_0 : \mu_1 = \mu_2 \quad \text{และ} \quad H_1 : \mu_1 \neq \mu_2$$

2. กำหนดระดับนัยสำคัญในการทดสอบ (α)
3. กำหนดค่าสถิติทดสอบ

$$z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

ถ้าไม่ทราบค่า σ_1^2 หรือ σ_2^2 ก็จะประมาณด้วยค่า S_1^2 และ S_2^2 ตามลำดับ

4. อ่านค่าวิกฤตจากตารางสถิติ

$$z_{crit} \leq -z_{\alpha/2} \quad \text{และ} \quad z_{crit} \geq z_{\alpha/2}$$

5. สรุปผลการทดสอบสมมติฐาน

หากค่าสถิติทดสอบที่คำนวณได้อยู่ในบริเวณวิกฤต จะปฏิเสธสมมติฐานหลัก นั่นคือค่าเฉลี่ยของประชากรทั้งสองกลุ่มจะมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับนัยสำคัญ α

เมื่อพิจารณาการคำนวณค่าสถิติทดสอบ หากขนาดความแตกต่างของข้อมูลตัวอย่างทั้งสองกลุ่ม ($\bar{x}_1 - \bar{x}_2$) มีขนาดใหญ่ ผลการทดสอบสมมติฐานควรจะแสดงความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่เมื่อพิจารณาปัจจัยที่มีผลต่อค่าสถิติทดสอบนอกจากขนาดของความแตกต่างของ

ข้อมูลตัวอย่างทั้งสองกลุ่มแล้วยังมีส่วนเบี่ยงเบนมาตรฐาน (σ) และจำนวนขนาดตัวอย่าง (n) อีกด้วย ดังนั้นถึงแม้ว่าขนาดความแตกต่างของข้อมูลตัวอย่างทั้งสองกลุ่มมีขนาดใหญ่ ค่าสถิติทดสอบที่คำนวณได้อาจจะไม่มีขนาดใหญ่เพียงพอที่จะปฏิเสธสมมติฐานหลักได้ เนื่องจากต้องคำนึงถึงความคลาดเคลื่อนมาตรฐานของข้อมูลประกอบด้วย

3. ขนาดความแตกต่างมาตรฐานของข้อมูลสองกลุ่มของ Cohen

Jacob Cohen [อ้างอิงใน 4] ได้เสนอแนวคิดในการศึกษาความแตกต่างของข้อมูล โดยพิจารณาจากขนาดของความแตกต่างมาตรฐานของข้อมูลที่ศึกษา โดยที่ไม่ได้คำนึงถึงจำนวนขนาดตัวอย่าง และไม่ต้องคำนึงถึงข้อสมมติเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติและความเท่ากันของความแปรปรวนของประชากรทั้งสองกลุ่ม ในกรณีที่ต้องการเปรียบเทียบความแตกต่างของประชากรสองกลุ่ม สามารถคำนวณค่าความแตกต่างมาตรฐานของ Cohen (d) ได้จากขนาดความแตกต่างของข้อมูลตัวอย่างทั้งสองกลุ่ม ($\bar{x}_1 - \bar{x}_2$) และส่วนเบี่ยงเบนมาตรฐานของประชากรทั้งสองกลุ่ม (σ) เท่านั้น

การคำนวณค่าความแตกต่างมาตรฐานของ Cohen (d) สามารถคำนวณได้ดังนี้ [5]

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sigma}$$

ถ้าไม่ทราบค่า σ สามารถประมาณด้วยค่า S_{Pooled} โดยที่

$$S_{Pooled} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}}$$

นอกจากนี้ยังสามารถคำนวณค่าความแตกต่างมาตรฐาน (d) จากค่าสถิติทดสอบ Z_{cal} ที่ได้จากการทดสอบความแตกต่างอย่างมีนัยสำคัญทางสถิติ โดยที่

$$d = Z_{cal} \sqrt{\left(\frac{n_1 + n_2}{n_1 n_2}\right) \left(\frac{n_1 + n_2}{n_1 + n_2 - 2}\right)}$$

เมื่อสามารถคำนวณค่าความแตกต่างมาตรฐานของ Cohen (d) ได้แล้วนั้น ในการพิจารณาว่าข้อมูลทั้งสองกลุ่มมีความแตกต่างกันอย่างไรนั้นสามารถพิจารณาได้จากค่าสัมประสิทธิ์สหสัมพันธ์ (r) โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์และค่าความแตกต่างมาตรฐานของ Cohen มีความสัมพันธ์กันดังนี้ [6]

$$r = \left(\frac{d^2}{d^2 + \frac{1}{p(1-p)}} \right)^{\frac{1}{2}}$$

$$\text{เมื่อ } p = \frac{n_1}{n_1 + n_2}$$

$$\text{และ } r = \frac{d}{\sqrt{d^2 + 4}} \quad \text{เมื่อ } p = \frac{1}{2}$$

ซึ่งเกณฑ์การสรุปผลเกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์นั้นหากมีค่าตั้งแต่ 0.5 จะสรุปว่าข้อมูลทั้งสองกลุ่มมีความสัมพันธ์กันมาก หากมีค่าตั้งแต่ 0.3 ถึง 0.5 แสดงว่าข้อมูลทั้งสองมีความสัมพันธ์กันปานกลาง และหากมีค่าตั้งแต่ 0.1 ถึง 0.3 แสดงว่าข้อมูลทั้งสองมีความสัมพันธ์กันน้อย [7] ดังนั้นจากความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์และค่าความแตกต่างมาตรฐานของ Cohen จึงสามารถสรุปผลเกี่ยวกับความแตกต่างของข้อมูลทั้งสองกลุ่มได้ดังตารางที่ 1

ตารางที่ 1 ความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์และค่าความแตกต่างมาตรฐานของ Cohen และผลสรุป

| สัมประสิทธิ์สหสัมพันธ์ (r) | ความแตกต่างมาตรฐานของ Cohen (d) | ความแตกต่างของข้อมูลทั้งสองกลุ่ม |
|--------------------------------|---------------------------------|----------------------------------|
| 0.1 | 0.20 | น้อย |
| 0.3 | 0.63 | ปานกลาง |
| 0.5 | 1.15 | มาก |

จากการพิจารณาความแตกต่างของประชากรสองกลุ่มด้วยแนวคิดของ Cohen ปัจจัยที่มีผลต่อค่าความแตกต่างมาตรฐาน (d) คือ ขนาดของความแตกต่างของข้อมูลตัวอย่างทั้งสองกลุ่ม และส่วนเบี่ยงเบนมาตรฐานเท่านั้น โดยที่ขนาดความแตกต่างของข้อมูลตัวอย่างทั้งสองกลุ่ม มีขนาด 0.20 เท่าของส่วนเบี่ยงเบนมาตรฐาน จะสรุปว่าข้อมูลทั้งสองกลุ่มนั้นมีความแตกต่างกันน้อย ถ้าขนาดความแตกต่างของข้อมูลตัวอย่างทั้งสองกลุ่ม มีขนาด 0.63 เท่าของส่วนเบี่ยงเบนมาตรฐาน สรุปว่ามีความแตกต่างกันปานกลาง และถ้าขนาดความแตกต่างของข้อมูลตัวอย่างทั้งสองกลุ่ม มีขนาด 1.15 เท่าของส่วนเบี่ยงเบนมาตรฐาน จะสรุปว่าข้อมูลทั้งสองกลุ่มนั้นมีความแตกต่างกันมาก

4. ตัวอย่างการวิเคราะห์ข้อมูล

ข้อมูลที่ใช้ในการวิเคราะห์คือข้อมูลจาก Journal of Statistics Education เป็นข้อมูลอุณหภูมิจนในร่างกาย (ฟาเรนไฮต์) ของเพศชายและหญิงจำนวน 130 คน แบ่งเป็นชายและหญิงจำนวนกลุ่มละ 65 คน ซึ่งข้อมูลทั้งสองกลุ่มมีการแจกแจงแบบปกติ [8]

นำข้อมูลดังกล่าวมาศึกษาความแตกต่างของอุณหภูมิในร่างกายเพศชายและหญิงว่ามีความแตกต่างกันหรือไม่ ผลการวิเคราะห์สามารถแสดงได้ดังตารางที่ 2

ตารางที่ 2 ผลการวิเคราะห์ความแตกต่างของอุณหภูมิในร่างกายเพศชายและหญิง

| กลุ่ม | อุณหภูมิเฉลี่ย | ส่วนเบี่ยงเบนมาตรฐาน | สถิติทดสอบ | P-Value |
|---------|----------------|----------------------|------------|---------|
| เพศหญิง | 98.393 | 0.743 | 2.285 | 0.024 |
| เพศชาย | 98.104 | 0.699 | | |

เมื่อกำหนดให้ μ_1 แทนอุณหภูมิเฉลี่ยในร่างกายเพศหญิง และ μ_2 แทนอุณหภูมิเฉลี่ยในร่างกายเพศชาย การทดสอบความแตกต่างอย่างมีนัยสำคัญทางสถิติของข้อมูลทั้งสองกลุ่ม ที่ระดับนัยสำคัญ 0.05 จะมีวิธีการดังนี้ คือ

1. กำหนดสมมติฐานของการทดสอบ

$$H_0 : \mu_1 = \mu_2 \quad \text{และ} \quad H_1 : \mu_1 \neq \mu_2$$

2. ระดับนัยสำคัญในการทดสอบคือ 0.05

3. ค่าสถิติทดสอบมีค่าเท่ากับ

$$z_{cal} = \frac{98.393 - 98.104}{\sqrt{\frac{(0.743)^2}{65} + \frac{(0.699)^2}{65}}} = 2.285$$

4. ค่าวิกฤตจากการเปิดตารางคือ

$$z_{crit} \leq -1.96 \quad \text{และ} \quad z_{crit} \geq 1.96$$

5. การสรุปผลที่ได้คือ เนื่องจากค่าสถิติทดสอบที่คำนวณได้มีค่าอยู่ในบริเวณวิกฤตจึงปฏิเสธสมมติฐานหลัก นั่นคือ เพศชายและเพศหญิงมีอุณหภูมิในร่างกายเฉลี่ยแตกต่างกัน

การปฏิเสธสมมติฐานหลักแสดงว่าข้อมูลทั้งสองกลุ่มมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ความแตกต่างนี้จะสามารถสรุปได้หรือไม่ว่าอุณหภูมิในร่างกายของเพศชายและเพศหญิงมีความแตกต่าง

กันจริง ทั้งนี้เมื่อพิจารณาอุณหภูมิเฉลี่ยของตัวอย่างทั้งสองกลุ่มจะมีค่าใกล้เคียงกันมากคือ เพศชายมีอุณหภูมิเฉลี่ย 98.104 ฟาเรนไฮต์ และเพศหญิงมีอุณหภูมิเฉลี่ย 98.393 ฟาเรนไฮต์

ผลสรุปที่ได้จากการทดสอบความมีนัยสำคัญทางสถิติจะทำให้เกิดความไม่มั่นใจในการนำผลสรุปไปใช้งานได้ ในกรณีนี้หากพิจารณาขนาดของความแตกต่างมาตรฐานของ Cohen (d) มีค่าเท่ากับ

$$d = \frac{98.393 - 98.104}{\sqrt{\frac{(64)(0.743)^2 + (64)(0.699)^2}{(65 + 65 - 2)}}} = 0.400$$

และสามารถคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ (r) ได้เท่ากับ

$$r = \frac{0.40}{\sqrt{(0.40)^2 + 4}} = 0.196$$

ผลสรุปที่ได้จากขนาดความแตกต่างมาตรฐานของ Cohen นี้แสดงว่าอุณหภูมิเฉลี่ยของเพศชายและเพศหญิงมีความแตกต่างกันในระดับที่น้อยถึงปานกลางเท่านั้น

จากข้อมูลตัวอย่างข้างต้นจะเห็นได้ว่าผลการทดสอบสมมติฐานสรุปได้ว่าอุณหภูมิของเพศชายและหญิงมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ถึงแม้ว่าขนาดของความแตกต่างของอุณหภูมิเฉลี่ยเพียง 0.289 ฟาเรนไฮต์เท่านั้น การที่ศึกษาค่าความแตกต่างมาตรฐานของ Cohen ประกอบร่วมด้วยจะช่วยทำให้ยืนยันผลการศึกษาค่าได้ดียิ่งขึ้น กล่าวคือถึงแม้ค่าเฉลี่ยของอุณหภูมิจะแตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่ขนาดความแตกต่างนี้อยู่ในระดับที่น้อยถึงปานกลางเท่านั้น ทำให้ในการสรุปผลการศึกษาก็อาจจะไม่สามารถยืนยันผลที่ได้จากการ

ทดสอบสมมติฐานเพียงอย่างเดียว อาจจำเป็นที่จะต้องศึกษาเพิ่มเติมเพื่อยืนยันผลการทดลองอีกครั้งหนึ่ง เพื่อให้ได้ข้อสรุปที่มีความน่าเชื่อถือในการทดลอง

5. บทสรุป

การเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของประชากรสองกลุ่ม การไม่ปฏิเสธสมมติฐานหลักของการทดสอบสมมติฐานทางสถิติไม่ได้เป็นการยืนยันว่าประชากรทั้งสองกลุ่มนั้นไม่แตกต่างกัน เพราะว่าปัจจัยที่มีผลต่อค่าสถิติทดสอบไม่ได้มีเพียงขนาดของความแตกต่างของข้อมูลตัวอย่างทั้งสองกลุ่มเท่านั้น แต่ยังรวมถึงความคลาดเคลื่อนมาตรฐานของข้อมูลอีกด้วย

ขณะที่การเปรียบเทียบความแตกต่างของประชากรสองกลุ่มด้วยการพิจารณาขนาดความแตกต่างมาตรฐานของข้อมูลสองกลุ่มของ Cohen จะพิจารณาขนาดของความแตกต่างของข้อมูลตัวอย่างทั้งสองกลุ่มและส่วนเบี่ยงเบนมาตรฐานเท่านั้น ไม่ได้คำนึงถึงจำนวนขนาดตัวอย่างมาเกี่ยวข้อง ความแตกต่างที่เกิดขึ้นจึงเกิดจากความแตกต่างของข้อมูลอย่างแท้จริง

ดังนั้นในการศึกษาทดลองใดๆ เมื่อทำการทดสอบสมมติฐานเพื่อแสดงความแตกต่างระหว่างค่าเฉลี่ยของประชากรสองกลุ่ม หากพบว่าประชากรทั้งสองกลุ่มมีหรือไม่มี ความแตกต่างกันอย่างมีนัยสำคัญทางสถิติก็ตาม ควรที่จะรายงานและสรุปผลของขนาดความแตกต่างมาตรฐานของข้อมูลสองกลุ่มของ Cohen ร่วมด้วย ทั้งนี้เพื่อให้ข้อมูลประกอบที่ครบถ้วนแก่ผู้อ่านในการตัดสินใจเชื่อในผลสรุปของการศึกษาที่ได้

6. เอกสารอ้างอิง

- [1] Kirk, E. R., Promoting Good Statistical Practices: Some Suggestions, Educational and Psychological Measurement, Vol. 61; pp. 213-218, 2001.
- [2] Gliner, A. J., Leech, L. N. and Morgan, A. G., Problems With Null Hypothesis Significance Test: What Do the Textbooks Say?, The Journal of Experimental Education, Vol. 71; pp. 83-92, 2002.
- [3] Mecklin, J. C., A Comparison of Equivalence Testing in Combination with Hypothesis Testing and Effect Size, Journal of Modern Applied Statistical Methods, Vol. 2; pp. 329-340, 2003.
- [4] Kromrey, D. J., Hogarty, Y. K., Ferron, M. J., Hines, V. C. and Hess, R. M., Robustness in Meta-Analysis: An Empirical Comparison of Point and Interval Estimates of Standardized Mean Differences and Cliff's Delta, Joint Statistical Meetings., 2005.
- [5] Thalheimer, W and Cook, S., How to Calculate Effect Sizes from Published Research Articles: A Simplified Methodology, http://www.work-learning.com/white_papers/effect_sizes/Effect_Sizes_pdf5.pdf (December, 21, 2008).
- [6] DeCoster, J., Meta-Analysis Notes. <http://www.bama.ua.edu/~jamie/meta> (December, 21, 2008).

- [7] Cohen, J., Statistical Power Analysis for the Behavioral Sciences, 2nd ed., Lawrence Erlbaum, New Jersey, 1988.
- [8] Shoemaker L.A., What's Normal? –Temperature, Gender, and Heart Rate, Journal of Statistics Education, Vol. 4; pp. 350-354, 1996.