อำนาจของการทดสอบทางสถิติ

ความหมาย

อำนาจของการทดสอบทางสถิติ (Power of Statistical Test) หมายถึง ความน่าจะเป็นของการปฏิเสธสมมติฐานกลาง (null hypothesis, H_0) ที่เป็นเท็จได้อย่างถูกต้องเมื่อทำการวิเคราะห์ด้วยวิธีการทางสถิติใด ๆ อำนาจของการทดสอบทางสถิติสามารถเขียนอธิบายให้อยู่ในรูปทั่วไปได้ดังนี้

อำนาจของการทดสอบทางสถิติ = $1 - \beta$ เมื่อ β แทน ความน่าจะเป็นของการไม่ปฏิเสธสมมติฐานกลาง เมื่อสมมุติฐานกลางนั้นเป็นเท็จ หรือ เรียกอีกอย่างหนึ่งว่าความคลาดเคลื่อนแบบที่ 2 (type II error)

ความเป็นมา

ในอดีตความสนใจของนักวิจัยทางพฤติกรรมศาสตร์ที่มีต่ออำนาจของการทดสอบทางสถิติมีน้อยมาก เมื่อเปรียบเทียบกับความสนใจและความตระหนักที่มีต่อ α หรือความน่าจะเป็นของการปฏิเสธสมมติฐานกลางที่เป็นจริง ซึ่งนักวิจัยนิยมเรียกว่า "ความคลาดเคลื่อนแบบที่ 1" (type I error) จนกระทั่งในปี ค.ศ. 1960 จาคอบ โคเฮน (Jacob Cohen) ศาสตราจารย์ด้านจิตวิทยาเชิงปริมาณ ณ มหาวิทยาลัยแห่งมลรัฐนิวยอร์ก (New York University) สหรัฐอเมริกา ได้เสนอแนวคิดและวิธีการวิเคราะห์อำนาจของการทดสอบทางสถิติขึ้น และในปี ค.ศ.1962 ผลงานของเขาได้รับการตีพิมพ์เผยแพร่อย่างเป็นทางการลงในวารสาร Journal of Abnormal and Social Psychology โดยมีรากฐานแนวคิดมาจาก "การทดสอบสมมุติฐานทางสถิติ" ตามวิธีการของเนแมนและเพียร์สัน (J. Neyman & E.S.Pearson) ที่อิงทฤษฎีการอนุมานทางสถิติ ซึ่งมีสาระสำคัญตรงข้ามกับแนวคิด "การทดสอบนัยสำคัญทางสถิติ" ที่เน้นทดสอบสมมติฐานกลางเป็นหลักตามวิธีดั้งเดิมของฟิชเชอร์ (R.A. Fisher) ซึ่งมีอิทธิพลอย่างมากต่อแนวคิดความเชื่อของนักวิจัยทางพฤติกรรมศาสตร์นับตั้งแต่เวลานั้นจวบจนถึงปัจจุบันนี้ ผลงานเกี่ยวกับอำนาจของการทดสอบทางสถิติของจาคอบ โคเฮน มาได้รับความสนใจอีกครั้ง เมื่อหนังสือตำราทางสถิติของเขา คือ Statistical Power Analysis of the Behavioral Science ได้รับการตีพิมพ์เผยแพร่ในปี ค.ศ. 1969 และตีพิมพ์ครั้งที่ 2 ในปี ค.ศ. 1988 ปัจจุบันนี้แนวคิดของจาคอบ โคเฮน ในเรื่องอำนาจการทดสอบทางสถิติได้รับความสนใจมากขึ้นเรื่อย ๆ ดังจะเห็นได้จากผลงานทางวิชาการของเขาที่พิมพ์เผยแพร่ไปแล้วได้รับการอ้างอิงบ่อยครั้งมากจากบรรดานักวิชาการด้านสถิติและนักวิจัยทางพฤติกรรมศาสตร์ นอกจากนี้ หนังสือตำราและบทความทางวิชาการ รวมทั้งโปรแกรมวิเคราะห์ข้อมูลทางด้านสถิติที่มีชื่อเสียงจำนวนมากในปัจจุบันนี้ได้กล่าวถึงหรือบรรจุเนื้อหาสาระเกี่ยวกับอำนาจของการทดสอบตามแนวคิดของเขาไว้ด้วยเสมอ ๆ

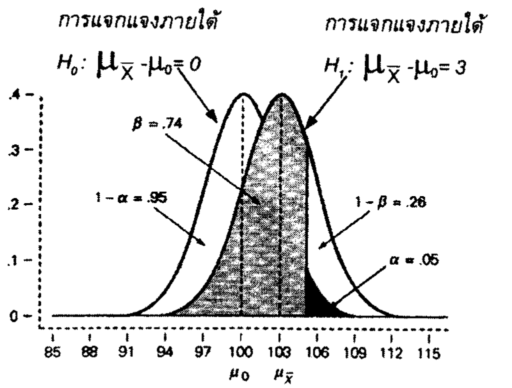
ปัจจัยที่มีอิทธิพลต่ออำนาจของการทดสอบทางสถิติ

อำนาจของการทดสอบด้วยวิธีการทางสถิติวิเคราะห์ใด ๆ ขึ้นอยู่กับปัจจัยที่สำคัญดังนี้

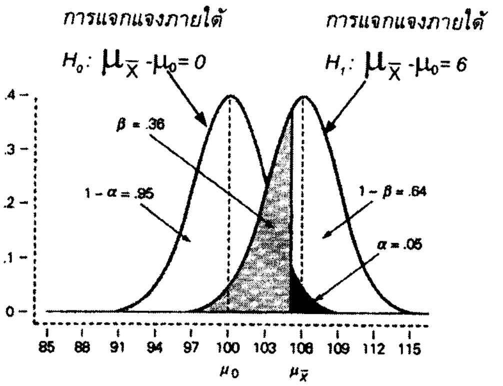
1. ขนาดความแตกต่างระหว่างค่าเฉลี่ยของประชากรตามสมมติฐาน ($\mu_{\bar{X}}$) และค่าเฉลี่ยประชากรที่แท้จริง (μ_0) มีผลต่ออำนาจของการทดสอบทางสถิติ ซึ่งสามารถอธิบายได้ดังแผนภาพ 1(ก) เมื่อกำหนดให้การแจกแจงค่าเฉลี่ยของกลุ่มตัวอย่างภายใต้สมมติฐานกลางที่ว่า $\mu_{\bar{X}} - \mu_0 = 0$

เป็นจริง และการแจกแจงค่าเฉลี่ยของกลุ่มตัวอย่าง เกิดขึ้นภายใต้สมมติฐานทางเลือก (alternative hypothesis, H_1) ที่ว่า $\mu_{\bar{X}} - \mu_0 = 3$ เป็นจริง จากแผนภาพ 1(ก) จะเห็นว่าเมื่อพิจารณาการแจกแจงทั้ง 2 แบบ ในลักษณะซ้อนทับกัน โดยการแจกแจงค่าเฉลี่ยของกลุ่มตัวอย่างภายใต้สมมติฐานกลางเป็นจริงมี $\mu_0 = 100$ และมีพื้นที่สีดำบ่งชี้ค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนแบบที่ 1 (α) หรือ การปฏิเสธสมมติฐานกลางที่เป็นจริง ซึ่งเป็นที่รู้จักและนิยมเรียกกันในหมู่นักวิชาการหรือผู้สนใจทางสถิติว่า "ระดับนัยสำคัญทางสถิติ" ในขณะที่การแจกแจงค่าเฉลี่ยของกลุ่มตัวอย่างภายใต้สมมติฐานทางเลือกที่เป็นจริง มี $\mu_{\bar{X}} = 103$ โดยพื้นที่สีเทาดังที่แสดงไว้ คือ β มีค่าเท่ากับ .74 ส่วนพื้นที่ทั้งหมดที่อยู่ทางด้านขวาพื้นที่สีเทาของการแจกแจงนี้ คือ $1 - \beta = .26$ แทนความน่าจะเป็นของการปฏิเสธ $H_0: \mu_{\bar{X}} - \mu_0 = 0$ ที่เป็นเท็จ (เมื่อ $H_1: \mu_{\bar{X}} - \mu_0 = 3$ เป็นจริง) ซึ่งก็คือ อำนาจของการทดสอบสมมุติฐานกลางดังกล่าว

อย่างไรก็ตาม จะพบว่าในกรณีที่การแจกแจงค่าเฉลี่ยของกลุ่มตัวอย่างภายใต้สมมติฐานทางเลือกเป็นจริง เมื่อ $\mu_{\bar{X}} = 106$ ดังที่แสดงในแผนภาพ 1(ข) จะพบว่า ความน่าจะเป็นของการปฏิเสธ $H_0: \mu_{\bar{X}} - \mu_0 = 0$ ที่เป็นเท็จ (เมื่อ $H_1: \mu_{\bar{X}} - \mu_0 = 6$ เป็นจริง) มีค่าเท่ากับ .64 หรือ กล่าวอีกนัยหนึ่ง คือ อำนาจของการทดสอบสมมุติฐานกลางดังกล่าวที่เกิดขึ้นในกรณีนี้มีค่าสูงกว่าในกรณีแรก ดังนั้น จึงสรุปได้ว่าเมื่อขนาดตัวอย่าง (n) ความแปรปรวนของประชากร (σ^2) และระดับนัยสำคัญทางสถิติ (α) คงที่แล้ว ขนาดของ $\mu_{\bar{X}} - \mu_0$ มีความสัมพันธ์กับอำนาจของการทดสอบทางสถิติ ซึ่งค่าขนาด $\mu_{\bar{X}} - \mu_0$ ภายใต้การแจกแจงทั้งสองแบบนี้ มักนิยมแสดงอยู่ในหน่วยมาตรฐาน คือ นำมาหารด้วย σ หรือ ส่วนเบี่ยงเบนมาตรฐานภายในกลุ่มรวม (common within-group $\sigma \cong \sqrt{MS_W}$) ซึ่งแท้ที่จริงก็คือ ค่า $d = |\mu_{\bar{X}} - \mu_0| / \sigma$ หรือค่าขนาดอิทธิพล (effect size) ที่รู้จักและนิยมใช้กันอย่างกว้างขวางในการวิจัยเชิงทดลองและการวิเคราะห์อภิมาน (meta-analysis)



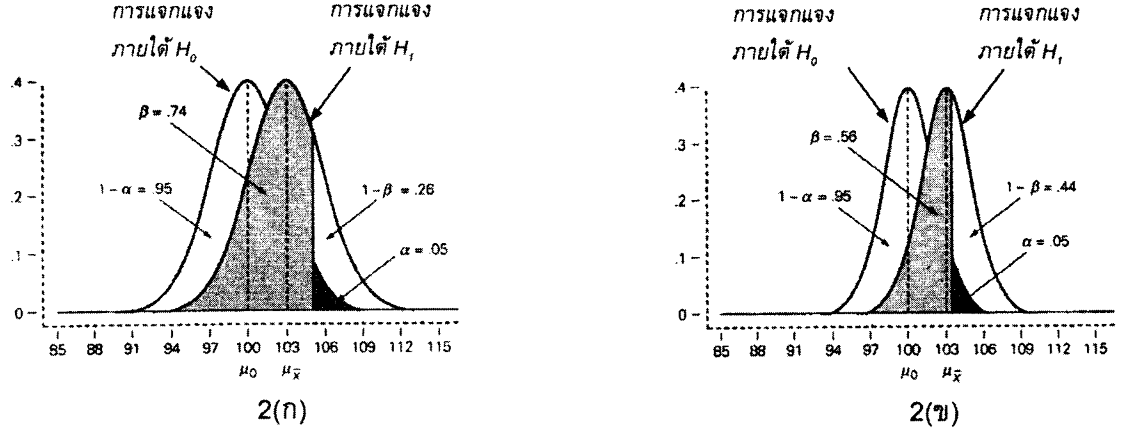
1(ก)



1(ข)

แผนภาพ 1: *อิทธิพลของความแตกต่างระหว่าง μ_0 และ $\mu_{\bar{X}}$ ต่อค่าอำนาจของการทดสอบ*

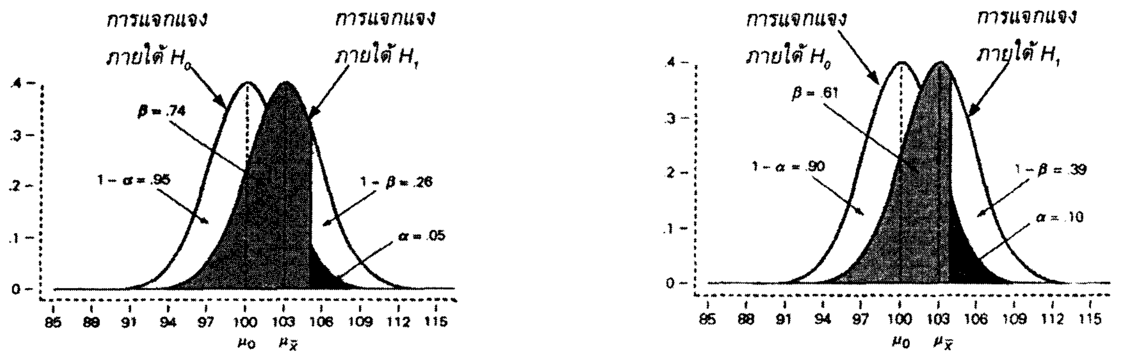
2. ขนาดกลุ่มตัวอย่าง กล่าวโดยหลักการทางสถิติแล้ว ขนาดกลุ่มตัวอย่างมีความสำคัญในการประมาณค่าพารามิเตอร์ของประชากร กล่าวคือ เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น จะมีความแม่นยำในการประมาณค่าจะเพิ่มขึ้นตามไปด้วยปรากฏการณ์ดังกล่าวนี้อธิบายได้ด้วยทฤษฎีลิมิตสู่ส่วนกลาง (central limit theorem) ที่ระบุว่า ถ้ากลุ่มตัวอย่างได้มาจากการสุ่มประชากรที่มีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ เมื่อเพิ่มกลุ่มตัวอย่าง n ให้มีขนาดมากขึ้นแล้ว การแจกแจงของค่าเฉลี่ยของกลุ่มตัวอย่าง ($\bar{x}$) จะมีลักษณะเข้าใกล้การแจกแจงแบบปกติที่มีค่าเฉลี่ย μ และมีส่วนเบี่ยงเบนมาตรฐาน (ความคลาดเคลื่อนมาตรฐาน) $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ จากสมการนี้ จะพบว่าเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ค่าความคลาดเคลื่อนมาตรฐานจะมีระดับต่ำ หรือมีความแม่นยำของการประมาณค่าพารามิเตอร์น้อย เป็นผลให้มีพื้นที่ความน่าจะเป็นของการแจกแจงภายใต้ H_0 และ H_1 เป็นจริง ซ้อนทับกันน้อยลง ซึ่งเป็นผลให้อำนาจของการทดสอบเพิ่มขึ้น เมื่อเปรียบเทียบกับอำนาจของการทดสอบในกรณีที่มีความคลาดเคลื่อนมาตรฐานระดับสูง หรือมีความแม่นยำของการประมาณค่าพารามิเตอร์มาก อันเนื่องมาจากการมีตัวอย่างขนาดเล็ก ดังรายละเอียดที่แสดงในแผนภาพ 2(ข) เมื่อเปรียบเทียบกับแผนภาพ 2(ก)



แผนภาพ 2 : *อิทธิพลของขนาดกลุ่มตัวอย่างต่ออำนาจของการทดสอบ*

ถ้าพิจารณาเฉพาะปัจจัยด้านความแปรปรวนของประชากรที่ศึกษาวิจัยต่ออำนาจของการทดสอบ จะพบว่า ถ้าความแปรปรวนของประชากรมีค่ามาก อำนาจของการทดสอบจะมีค่าในระดับต่ำ ในทางตรงกันข้ามถ้าความแปรปรวนของประชากรมีค่าน้อย อำนาจของการทดสอบจะมีค่าในระดับสูง ดังจะเห็นได้จากตัวอย่างที่แสดงในแผนภาพ 2(ก) ค่า $1 - \beta = .26$ ในขณะที่แผนภาพ 2(ข) $1 - \beta = .44$ ทั้งนี้เนื่องจากระดับการซ้อนทับกันของการแจกแจงทั้งสองแบบในกรณีความแปรปรวนของประชากรมีมาก ผลที่ตามมาทำให้อำนาจของการทดสอบมีค่าอยู่ในระดับต่ำแต่โอกาสการซ้อนทับกันดังกล่าวนี้จะมีน้อย เมื่อความแปรปรวนของประชากรมีค่าน้อย ซึ่งเป็นผลให้อำนาจของการทดสอบอยู่ในระดับสูง ดังที่แสดงไว้ในแผนภาพ 2(ก) และ 2(ข) ตามลำดับ

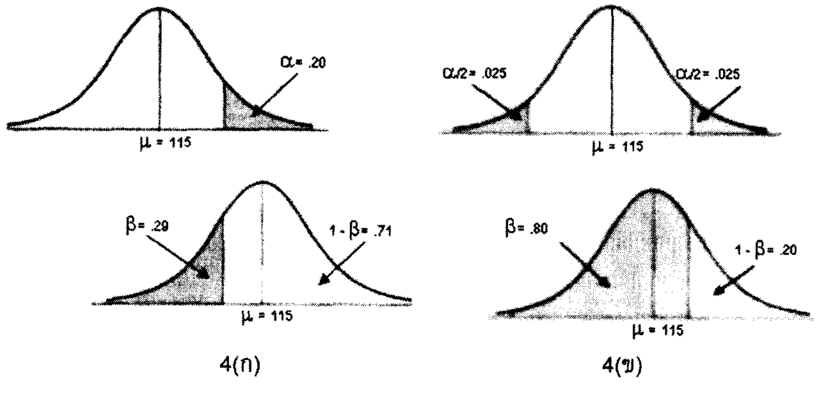
3. ระดับนัยสำคัญทางสถิติ ระดับนัยสำคัญทางสถิติหรือความคลาดเคลื่อนแบบที่ 1 (α) มีผลต่อความน่าจะเป็นในการปฏิเสธสมมติฐานกลางที่เป็นจริง ดังที่แสดงในแผนภาพ 3(ก) ถ้าระดับนัยสำคัญทางสถิติมีระดับต่ำ จะทำให้ β หรือความคลาดเคลื่อนแบบที่ 2 มีค่าสูงขึ้น เป็นผลทำให้อำนาจของการทดสอบทางสถิติลดลงในทางกลับกัน ถ้าเพิ่มระดับนัยสำคัญทางสถิติจะส่งผลให้ β มีค่าลดต่ำลง เป็นผลทำให้อำนาจของการทดสอบทางสถิติจะมีค่าเพิ่มสูงขึ้น ดังแผนภาพ 3(ข)



3(ก) 3(ข)

แผนภาพ 3 : *อิทธิพลจากระดับนัยสำคัญทางสถิติต่ออำนาจของการทดสอบ*

4. ทิศทางการทดสอบ การเลือกการทดสอบแบบมีทิศทาง (การทดสอบแบบทางเดียว) หรือไม่มีทิศทาง (การทดสอบแบบสองทาง) มีอิทธิพลต่ออำนาจของการทดสอบ การทดสอบแบบทางเดียวจะให้ค่าอำนาจของการทดสอบสูงกว่าการทดสอบแบบสองทาง เพราะการทดสอบแบบสองทางเขตวิกฤต (critical region) จะมีขนาดเป็นครึ่งหนึ่งของการทดสอบแบบทางเดียวโดยแต่ละส่วนอยู่ที่ด้านปลายของการแจกแจง ดังนั้น เมื่อใช้การทดสอบแบบสองทางจะทำให้ความน่าจะเป็นของการปฏิเสธสมมติฐานกลางมีระดับต่ำกว่าใช้การทดสอบแบบทางเดียว ดังจะเห็นได้อย่างชัดเจนจากแผนภาพ 4(ก) และ 4(ข) ที่แสดงค่าอำนาจของการทดสอบด้วยการทดสอบที (t-test) เมื่อกำหนดให้ $|\mu - \mu_0| = 2.5$, $\hat{\sigma} = 12.5$, n=60 และ $\alpha = .05$ โดยจะพบว่า เมื่อทำการทดสอบแบบทางเดียว ค่า $1 - \beta = .71$ แต่เมื่อทำการทดสอบแบบสองทาง $1 - \beta = .20$ เท่านั้น



4(ก) 4(ข)

แผนภาพ 4: *อิทธิพลของการกำหนดทิศทางการทดสอบต่ออำนาจของการทดสอบ*

การประมาณค่าอำนาจของการทดสอบ

การประมาณค่าอำนาจของการทดสอบกระทำได้ 2 แบบ คือ การประมาณค่าอำนาจของการทดสอบก่อนและหลังทำการทดลอง การประมาณค่าอำนาจของการทดสอบหลังทำการทดลองมีความสำคัญด้านการตีความผลการทดลอง นักวิจัยที่ไม่สนใจอำนาจของการทดสอบอาจจะแปลผลการทดลองที่พบว่าสิ่งทดลอง (treatment) ที่จัดกระทำขึ้นมีความแตกต่างอย่างไม่มีนัยสำคัญทางสถิติทั้งแท้ที่จริงแล้วสิ่งทดลองแตกต่างกัน ทั้งนี้เนื่องจากอำนาจของการทดสอบมีค่าอยู่ในระดับต่ำ ทำให้ไม่สามารถตรวจพบขนาดอิทธิพลหรือความแตกต่างระหว่างค่าเฉลี่ยของประชากรตามสมมติฐานและค่าเฉลี่ยประชากรที่แท้จริงได้ อำนาจของการทดสอบในระดับต่ำอาจเป็นผลมาจากตัวอย่างและอิทธิพลของสิ่งทดลองที่จัดกระทำขึ้นมีขนาดเล็ก ซึ่งการประมาณค่าอำนาจของการทดสอบหลังทำการทดลองสามารถกระทำได้หลายวิธีการ ได้แก่ การเปิดค่าจากตารางโดยตรง เช่น ตารางของโคเฮน ซึ่งมักนิยมใช้กันในหมู่นักสถิติและนักวิจัย การตรวจหาค่าจากแผนภูมิ การใช้โปรแกรม เช่น GPOWER และ PASS นอกจากนี้ ยังประมาณค่าโดยการคำนวณจากสูตร เช่น สูตรที่เสนอโดย โรเจอร์ เคิร์ค (Roger E. Kirk) ซึ่งพัฒนามาจากแนวความคิดของดิกซอนและแมสซี่ (Dixon & Massey) ที่เสนอไว้ในปี ค.ศ. 1957 สำหรับใช้คำนวณหาค่าอำนาจของการทดสอบทางสถิติด้วยวิธีการทดสอบที ส่วนสูตรที่ใช้คำนวณหาค่าอำนาจของการทดสอบด้วยวิธีการทางสถิติอื่นๆ ผู้อ่านที่สนใจสามารถศึกษาได้จากหนังสือตำราทางสถิติที่มีเนื้อหาครอบคลุมหัวข้อการวิเคราะห์อำนาจของการทดสอบ ในบทความนี้จะแสดงตัวอย่างเฉพาะการคำนวณหาอำนาจของการทดสอบทางสถิติด้วยการทดสอบที โดยใช้สูตรสำหรับในกรณีตัวอย่างสองกลุ่มที่เป็นอิสระจากกัน (two-sample t-test for independent samples) โดยมีจุดมุ่งหมายหลักเพื่อแสดงให้เห็นว่าภายใต้เงื่อนไขเดียวกัน อำนาจของการทดสอบทางสถิติที่ประมาณค่าด้วยวิธีการต่างๆ ที่จะกล่าวถึงต่อไปนี้จะมีค่าเท่ากัน

การใช้สูตรคำนวณ: สมมติให้ $\mu_0 = 115$, $\hat{\sigma} = 12.5$ และ $n = 60$ ในการคำนวณหาค่าอำนาจของการทดสอบทางสถิติด้วยการทดสอบที สำหรับกรณีตัวอย่างสองกลุ่ม จำเป็นจะต้องทราบความแตกต่างระหว่าง $\mu_{\bar{x}}$ และ μ_0 ในที่นี้กำหนดให้ $|\mu_{\bar{x}} - \mu_0| = 2.5$ สำหรับสูตรที่ใช้คำนวณหาค่าอำนาจของการทดสอบทางสถิติด้วยการทดสอบที สำหรับกรณีตัวอย่างสองกลุ่มที่ทดสอบแบบทางเดียว (Kirk, 1995: 62) มีดังนี้

$$z_\beta = \frac{d\,(n-1)\sqrt{2n}}{(n-1) + 1.21\,(z_\alpha - 1.06)} - Z_\alpha \quad \dots (1)$$

$$\text{เมื่อ} \quad d = \frac{|\mu_{\bar{x}} - \mu_0|}{\sigma} \quad \dots (2)$$

Z_α และ Z_β คือ คะแนนซีมาตรฐานที่เกิดจากความคลาดเคลื่อนแบบที่ 1 และที่ 2 ส่วนค่า 1.21 และ −1.06 เป็นค่าคงที่ ถ้าเป็นการทดสอบแบบสองทางจะแทนค่า Z_α ในสูตรที่ (1) ด้วย $Z_{\alpha/2}$ ซึ่งส่วนใหญ่ความน่าจะเป็นของการเกิดความคลาดเคลื่อนแบบที่ 1 มักกำหนดเป็น .05 ดังนั้น จะได้ค่าของ $Z_{.05}$ และ $Z_{.05/2}$ มีค่าเท่ากับ 1.645 และ 1.960 ตามลำดับ และถ้าเป็นการทดสอบสำหรับกรณีตัวอย่างกลุ่มเดียว (one-sample t-test) จะใช้ค่า $\sqrt{n}$ แทน $\sqrt{2n}$ ในสูตร (1) ที่ระบุข้างต้น เมื่อพิจารณาในสมการ (2) ค่าของ σ จะใช้ $\hat{\sigma}$ ซึ่งเป็นค่าประมาณพารามิเตอร์แทนในสูตรที่ (2) จะได้

$$d = \frac{2.5}{12.5} = 0.2$$

ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 2 ของคะแนนมาตรฐาน คือ

$$Z_\beta = \frac{0.2\ (60-1)\sqrt{2 \times 60}}{(60-1)+1.21\ (1.645-1.06)} - 1.645 = -0.556$$

เมื่อเปิดตารางโค้งปกติ (normal curve) จะได้พื้นที่ทางด้านขวาของ Z_β = −0.556 เท่ากับ .21+.50 ≅ .71 ดังนั้น อำนาจของการทดสอบที (power of t-test) มีค่าเท่ากับ 1 − β = 1 − .71 = .29

การเปิดหาค่าจากตาราง: นอกจากการใช้สูตรที่เคิร์คพัฒนาขึ้นตามแนวคิดของดิกซอนและแมสซี่สำหรับใช้คำนวณหาค่าประมาณอำนาจของการทดสอบทางสถิติข้างต้นแล้ว ยังสามารถประมาณค่าได้จากการเปิดตารางอำนาจของการทดสอบที่เสนอโดยจาคอบ โคเฮน ในที่นี้จะเริ่มจากประมาณค่าขนาดอิทธิพลโดยคำนวณจากสูตร (2) (ในทางปฏิบัติค่าขนาดอิทธิพลสามารถกำหนดได้จากประสบการณ์ของผู้วิจัยหรืองานวิจัยที่เกี่ยวข้องในอดีต) จากนั้นนำค่าขนาดอิทธิพลที่ได้ไปเปิดตารางเพื่อหาค่าอำนาจของการทดสอบทางสถิติจากตารางที่จาคอบ โคเฮน พัฒนาขึ้นจำนวนหลายตารางตามสถิติแต่ละวิธีการที่ใช้ สำหรับในบทความนี้ จะขอยกตัวอย่างเฉพาะตารางแสดงค่าอำนาจของการทดสอบทางสถิติด้วยวิธีการการทดสอบที ตารางแสดงค่าอำนาจของการทดสอบด้วยสถิติวิธีการนี้ มีข้อกำหนดว่าจะต้องเป็นการทดสอบที สำหรับกรณีตัวอย่างสองกลุ่ม โดยแต่ละกลุ่มจะต้องมีขนาดเท่ากันและเป็นอิสระจากกัน เช่น ถ้าใช้จำนวนที่กำหนดไว้ในตัวอย่างข้างต้น กล่าวคือ ขนาดอิทธิพลมีค่าเท่ากับ 0.2 และเมื่อนำไปเปิดตารางของจาคอบ โคเฮน จะได้ค่าอำนาจของการทดสอบทางสถิติด้วยการทดสอบที ณ α = .05 เมื่อ n = 60 โดยการทดสอบแบบทางเดียว เท่ากับ .29 ดังค่าที่ปรากฏในตาราง 1 ซึ่งจะเห็นว่ามีค่าเท่ากับที่ได้จากการคำนวณโดยใช้สูตร (1) ดังแสดงข้างต้น

ตาราง 1: อำนาจของการทดสอบทางสถิติ $\mu_1 = \mu_2$ ที่ α = .05 ด้วยวิธีการทดสอบที[1]

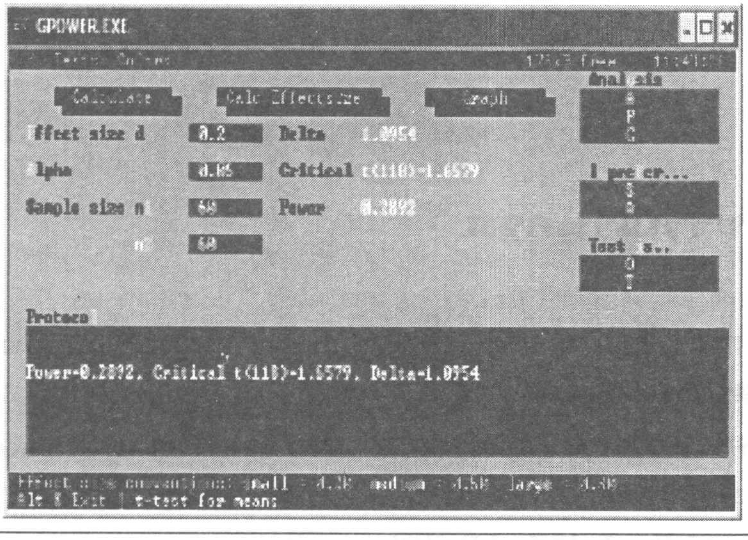
		d										
n	d_c	.10	.20	.30	.40	.50	.60	.70	.80	1.00	1.20	1.40
50	.33	12	26	44	63	80	91	97	99	*	*	*
52	.33	13	26	45	65	81	92	97	99			
54	.32	13	27	46	66	83	S93	98	99			
56	.31	13	28	47	68	84	93	98	99			
58	.31	13	28	49	69	85	94	98	*			
60	.30	13	29[2]	50	70	86	95	98				
64	.29	14	30	52	73	88	96	99				
68	.28	14	31	54	75	90	97	99				
72	.28	15	33	56	77	91	97	99				
76	.27	15	34	58	79	92	98	*				

หมายเหตุ

[1] ตัดตอนมาจาก Jacob Cohen. (1988: p. 31). Statistical Power Analysis of the Behavioral Science.

[2] ค่าอำนาจของการทดสอบ $\mu_1 = \mu_2$ ที่ α = .05 ด้วยวิธีการทดสอบที เมื่อ n = 60 มีค่าเท่ากับ 0.29

การใช้โปรแกรมคอมพิวเตอร์: ในปัจจุบันนี้มีโปรแกรมคอมพิวเตอร์หลายโปรแกรมที่ใช้ประมาณค่าอำนาจของการทดสอบทางสถิติ ซึ่งกระทำได้สะดวกและรวดเร็วกว่าการประมาณค่าโดยใช้วิธีการทั้งสองแบบที่กล่าวมาข้างต้น เช่นโปรแกรม GPOWER และ PASS ที่นักสถิติผู้มีชื่อเสียงจำนวนหลายคน เช่น Myer & Well (2003) และ Stevens (2002) แนะนำให้ใช้ ต่อไปนี้จะแสดงแผนภาพหน้าต่างหลัก (main window) ของโปรแกรม GPOWER หลังจากที่ใช้ประมาณค่าอำนาจของการทดสอบทีแบบตัวอย่างสองกลุ่มเป็นอิสระจากกัน ณ ระดับนัยสำคัญ (α) = .05 ภายใต้เงื่อนไขมีขนาดอิทธิพล (d) = 0.20 และ ขนาดกลุ่มตัวอย่าง (n) = 60 คน โดยใช้การทดสอบแบบทางเดียว จะพบว่า มีค่าเท่ากับ .29 ซึ่งเท่ากับค่าที่ได้จากการใช้สูตรคิดคำนวณ และที่เปิดหาค่าจากตารางของจาคอบ โคเฮน ดังที่แสดงมาข้างต้น



การใช้โปรแกรม GPOWER ประมาณค่าอำนาจของการทดสอบทางสถิติทีกระทำได้อย่างง่ายๆ เพียงใส่ค่า d = 0.20, α = .05, n_1 = n_2 = 60 และกำหนดให้การวิเคราะห์เป็นแบบภายหลังและทดสอบแบบทางเดียวลงในช่องรับข้อมูลบนหน้าต่างหลักนี้เท่านั้น

แผนภาพ 5: หน้าต่างหลักของโปรแกรม GPOWER ภายหลังประมาณค่าอำนาจของการทดสอบที เมื่อ α = .05, d = 0.20, และ n = 60

การประมาณค่าอำนาจของการทดสอบก่อนทำการทดลองต้องมีการประมาณค่าขนาดอิทธิพลโดยประมาณค่าจากการศึกษานำร่อง หรือการศึกษาทบทวนงานวิจัยที่เกี่ยวข้อง การประมาณค่าขนาดอิทธิพลโดยทั่วไปมักประมาณค่าโดยกำหนดเป็น 3 ขนาด คือ ขนาดเล็ก (d=.20) ขนาดกลาง (d=.50) และขนาดใหญ่ (d=.80) ตามเกณฑ์ของจาคอบ โคเฮน การประมาณค่าก่อนทำการทดลองใช้ประโยชน์ในการตัดสินใจกำหนดขนาดกลุ่มตัวอย่าง เช่น เมื่ออิทธิพลของสิ่งทดลองที่จัดกระทำขึ้นมีขนาดเล็ก ผู้วิจัยควรเพิ่มขนาดของกลุ่มตัวอย่างเพื่อให้มีอำนาจของการทดสอบพอเพียงสำหรับตรวจพบความแตกต่างอย่างมีนัยสำคัญ เมื่อพิจารณาอย่างถี่ถ้วนแล้วเห็นว่าขนาดอิทธิพลของสิ่งทดลองดังกล่าวมีความสำคัญในทางปฏิบัติตามทัศนะของผู้วิจัย รวมทั้งการยอมรับของนักวิชาการ นักวิจัยคนอื่นๆ ตลอดจนผู้อ่านหรือบุคคลผู้สนใจงานวิจัยนั้น

การใช้ประโยชน์จากอำนาจของการทดสอบทางสถิติ

อำนาจของการทดสอบทางสถิติที่ประมาณค่าได้ มีประโยชน์สำหรับใช้ในการออกแบบการวิจัย โดยเฉพาะอย่างยิ่ง การวิจัยเชิงทดลอง และยังมีประโยชน์สำหรับใช้ประเมินความสำเร็จของทำการวิจัย กล่าวอย่างเจาะจง คือ การใช้ประโยชน์

ประการแรกเป็นการใช้เพื่อตัดสินใจเกี่ยวกับการกำหนดขนาดตัวอย่างที่จำเป็นและระดับนัยสำคัญทางสถิติที่ใช้สำหรับทดสอบสมมุติฐาน ส่วนการใช้ประโยชน์ประการหลังเป็นการใช้เพื่อเป็นข้อมูลสำหรับประเมินโอกาสของความสำเร็จหรือความล้มเหลวของการทำวิจัย คือ การตรวจพบขนาดอิทธิพลของสิ่งทดลองที่ผู้วิจัยจัดกระทำหรือแทรกเสริมตามเงื่อนไขที่ได้กำหนดขึ้น ณ ระดับนัยสำคัญทางสถิติหนึ่งๆ ว่ามีหรือไม่

องอาจ นัยพัฒน์

บรรณานุกรม

องอาจ นัยพัฒน์. (2544). "อำนาจของการทดสอบทางสถิติ : ข้อควรคำนึงถึงสำหรับการกำหนดขนาดตัวอย่างเพื่อการวิจัย," **วารสารพฤติกรรมศาสตร์.** 7(1), 1 - 19.

Cohen, Jacob. (1988). **Statistical Power Analysis of the Behavioral Science.** 2nd ed., Hillsdale, NJ: Larence Erlbaum.

Cohen, Jacob. (1998). "A power primer," **Psychological Bulletin.** 112(1), 155-159.

Huberty, Carl J. (1987). "On statistical testing," **Educational Researcher.** 16(8), 4-9.

Kirk, Roger E. (1995). **Experimental Design: Procedures for the Behavioral Science.** 3rd ed. Pacific Grove, CA: Brooks/Cole.

Lipsey, Mark W. (1990). **Design sensitivity: Statistical power for experimental research.** Newbury Park, CA: Sage.

Myer, Jerome L.; & Well, Arnold D. (2003). **Research Design and Statistical Analysis.** 2nd ed. New Jersey: Lawrence Erlbaum.

Murphy, Kevin R., & Myors, B. (2004). **Statistical power analysis: a simple and general model for traditional and modern hypothesis tests.** 2nd ed. Mahwah, NJ: Lawrence Erlbaum.

Rosenthal, Robert; & Rosnow, Ralph L. (1991). **Essentials of Behavioral Research: Method and Data Analysis.** 2nd ed. New York: McGraw-Hill.

Runyon, Richard P., Haber, Audrey, Pittenger, David J., & Coleman, Kay A. (2000). **Fundamentals of Behavioral Statistics.** 9th ed. New York: McGraw-Hill.

Stevens, J. (2002). **Applied multivariate statistics for the social sciences.** 4th ed. Hillsdale, NJ: Lawrence Erlbaum.