ประเด็นเพิ่มเติมสำหรับการทดสอบสมมติฐาน

สถิติสำหรับจิตวิทยา 1

สันทัด พรประเสริฐมานิต

โครงร่างการนำเสนอ

- ขนาดอิทธิพล (Effect Size)
- ช่วงเชื่อมั่นของขนาดอิทธิพล (Confidence Interval of Effect Size)
- กำลังในการทดสอบทางสถิติ (Statistical Power)
- การกำหนดขนาดกลุ่มตัวอย่าง
- การกำหนดระดับนัยสำคัญทางสถิติ
- ข้อผิดพลาดที่มักเกิดขึ้นในการทดสอบสมมติฐาน
- การทดสอบความใกล้เคียง (Equivalence Testing)

ขนาดอิทธิพล

- ในครั้งที่ผ่านมา ได้กล่าวถึงการทดสอบสมมติฐาน ในกรณีที่เราเก็บข้อมูลจากคนเดียว และหลายคน
- ได้พูดถึง “นัยสำคัญทางสถิติ (Statistical Significance)” ว่าหมายถึง Null Hypothesis มีแนวโน้มไม่เป็นจริง
  - เช่น เขาไม่ใช่คนปกติ; เขามี IQ สูงกว่าปกติ
- แต่ “นัยสำคัญทางสถิติ” ไม่ได้กล่าวถึงความรุนแรงของความแตกต่าง
  - เช่น เขาไม่ใช่คนปกติ แล้วเขาไม่ปกติขนาดไหน
  - เขามี IQ สูงกว่าปกติมากน้อยเพียงใด
- สิ่งที่จะบอก คือ “ขนาดอิทธิพล (Effect Size)”

ขนาดอิทธิพล

- การทดสอบสมมติฐานสองกลุ่ม ได้ระดับนัยสำคัญเหมือนกัน

คอร์สฝึกอบรม A
ทำให้คนปกติ 10 คนมี
ค่าเฉลี่ย IQ เท่ากับ 110

$z = 2.11$ → $p = .017$

คอร์สฝึกอบรม B
ทำให้คนปกติ 40 คนมี
ค่าเฉลี่ย IQ เท่ากับ 105

$z = 2.11$ → $p = .017$

$M_A > M_B$ แต่ $p_A = p_B$

ตกลงคอร์ส
ไหนดีกว่ากัน

ขนาดอิทธิพล

- การทดสอบสมมติฐานสองกลุ่ม ได้ระดับนัยสำคัญเหมือนกัน

คอร์สฝึกอบรม A	คอร์สฝึกอบรม C
ทำให้คนปกติ 10 คนมี ค่าเฉลี่ย IQ เท่ากับ 110	ทำให้คนปกติ 40 คนมี ค่าเฉลี่ย IQ เท่ากับ 110
$z = 2.11$ → $p = .017$	$z = 4.22$ → $p = .000012$

$M_A = M_C$ แต่ $p_A > p_C$

ตกลงคอร์สไหนดีกว่ากัน

ขนาดอิทธิพล

- ค่าเฉลี่ยเป็นสิ่งที่บอกขนาดอิทธิพล เพราะจริงแล้วเราต้องการให้ค่าเฉลี่ยเปลี่ยนแปลง
  - เช่น พัฒนา IQ ก็ต้องการให้ค่า IQ (ค่าเฉลี่ยของกลุ่ม) สูงที่สุด
- ค่า *p* ไม่ได้บอกขนาดอิทธิพล
  - เช่น คอร์ส A และ คอร์ส C ทำให้ค่าเฉลี่ยของ IQ เพิ่มขึ้นเท่ากัน แต่ *p* แตกต่างกัน
  - เพราะเก็บจำนวนกลุ่มตัวอย่างแตกต่างกัน

ขนาดอิทธิพล

- คอร์สฝึกอบรม A ทำให้ค่า IQ เพิ่มขึ้นมากกว่าคนปกติ 10 คะแนน
- คอร์สฝึกอบรม B ทำให้ค่า IQ เพิ่มขึ้นมากกว่าคนปกติ 5 คะแนน
- คอร์สฝึกอบรม C ทำให้ค่า IQ เพิ่มขึ้นมากกว่าคนปกติ 10 คะแนน

ขนาดอิทธิพลที่คำนวณจากคะแนนดิบ

(Raw Score Effect Size)

อย่างไรก็ตาม การคำนวณด้วยคะแนนดิบ

อาจเปรียบระหว่างมาตรวัดที่แตกต่างกันไม่ได้

ขนาดอิทธิพล

ไม่ได้ เพราะการ
กระจายแตกต่างกัน

- เ

เปรียบเทียบคะแนนเลขที่เพิ่มขึ้นกับคะแนนอังกฤษที่เพิ่มขึ้นได้ไหม

โรงเรียนกวดวิชา A
สอนเลข ทำให้นักเรียน 100 คน
ได้คะแนนเลข 50 คะแนน

นักเรียนปกติ: $\mu = 45$, $\sigma = 10$

คะแนนเพิ่มขึ้น 5 คะแนน

โรงเรียนกวดวิชา B
สอนอังกฤษ ทำให้นักเรียน 100 คน
ได้คะแนนอังกฤษ 70 คะแนน

นักเรียนปกติ: $\mu = 65$, $\sigma = 20$

คะแนนเพิ่มขึ้น 5 คะแนน

ขนาดอิทธิพล

- วิธีการแก้ไข คือ ทำให้เป็นมาตรฐาน (Standardize) ด้วยความแตกต่างของค่าเฉลี่ยในหน่วยของส่วนเบี่ยงเบนมาตรฐาน
- จะเรียนค่านี้ว่า ขนาดอิทธิพลมาตรฐาน (Standardized Effect Size) หรือค่า *d* ของ Cohen (Cohen's *d*)

$$\delta = \frac{\mu_1 - \mu_0}{\sigma}$$

$$d = \frac{M_1 - \mu_0}{\sigma}$$

คือ ค่าเฉลี่ยของกลุ่มที่ได้รับอิทธิพล ลบด้วยกลุ่มที่ไม่ได้รับอิทธิพล หารด้วยส่วนเบี่ยงเบนมาตรฐานของกลุ่มที่ไม่ได้รับอิทธิพล

** ส่วนเบี่ยงเบนมาตรฐาน ไม่ใช่ Standard Error

ขนาดอิทธิพล

- เช

โรงเรียนกวดวิชา A มีขนาดอิทธิพลเป็นสองเท่าของโรงเรียน B

โรงเรียนกวดวิชา A	โรงเรียนกวดวิชา B
โรงเรียนกวดวิชา A สอนเลข ทำให้นักเรียน 100 คน ได้คะแนนเลข 50 คะแนน	โรงเรียนกวดวิชา B สอนอังกฤษ ทำให้นักเรียน 100 คน ได้คะแนนอังกฤษ 70 คะแนน
นักเรียนปกติ: $\mu = 45$, $\sigma = 10$	นักเรียนปกติ: $\mu = 65$, $\sigma = 20$
$d = (50 - 45)/10 = 0.5$	$d = (70 - 65)/20 = 0.25$

ขนาดอิทธิพล

- Cohen (1988) ตั้งเกณฑ์อย่างง่ายในการประเมินขนาดอิทธิพลในงานวิจัยทางจิตวิทยา

$d = 0.2$	น้อย	ต้องเพ่งดีๆ ถึงเห็นความแตกต่าง
$d = 0.5$	ปานกลาง	ความแตกต่างมองเห็นด้วยตาเปล่า
$d = 0.8$	มาก	มองผ่านๆ ก็เห็นความแตกต่าง

การหาช่วงเชื่อมั่นของขนาดอิทธิพล

- การหาช่วงเชื่อมั่นของขนาดอิทธิพล ว่าค่า Parameter ของขนาดอิทธิพลอยู่ในช่วงอะไร
- เช่น ในคอร์สการติวภาษาอังกฤษ มีขนาดอิทธิพลเท่ากับ 0.5 เมื่อเปรียบเทียบกับนักเรียนทั่วไป อยากทราบช่วงเชื่อมั่นระดับ .95 ว่าคอร์สนี้ทำให้นักเรียนเปลี่ยนแปลงอย่างไร

การหาช่วงเชื่อมั่นของขนาดอิทธิพล

- จากสูตร

$$\boxed{CI_{1-\alpha} \text{ of } \mu = M \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

- แปลงเทอมซ้ายให้เป็นขนาดอิทธิพล

1. ลบด้วย μ_0 ทั้งสองข้าง

$$CI_{1-\alpha} \text{ of } \mu - \mu_0 = M - \mu_0 \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

2. หารด้วย σ ทั้งสองข้าง

$$CI_{1-\alpha} \text{ of } \frac{\mu - \mu_0}{\sigma} = \frac{1}{\sigma}\left(M - \mu_0 \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

การหาช่วงเชื่อมั่นของขนาดอิทธิพล

- จากสูตร

$$CI_{1-\alpha} \text{ of } \mu = M \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- แปลงเทอมซ้ายให้เป็นขนาดอิทธิพล

3. เปลี่ยนเทอมซ้ายเป็น δ และกระจาย σ เข้าในสมการ

$$CI_{1-\alpha} \text{ of } \delta = \frac{M-\mu_0}{\sigma} \pm \frac{1}{\sigma}\left(z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

4. เปลี่ยนเทอมขวาเป็น *d* และตัด σ ที่เป็นเศษและส่วนออกจากสมการ

$$CI_{1-\alpha} \text{ of } \delta = d \pm \frac{z_{\alpha/2}}{\sqrt{n}}$$

การหาช่วงเชื่อมั่นของขนาดอิทธิพล

- เช่น ในคอร์สกวดวิชาภาษาอังกฤษ จำนวน 100 คน ได้ค่าเฉลี่ยเท่ากับ 70 คะแนน ขณะที่นักเรียนทั่วไปได้คะแนน 60 คะแนน ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 20 คะแนน
- ทำให้มีขนาดอิทธิพลเท่ากับ 0.5 อยากทราบช่วงเชื่อมั่นระดับ .95 ว่าคอร์สนี้ทำให้นักเรียนเปลี่ยนแปลงอย่างไร

การหาช่วงเชื่อมั่นของขนาดอิทธิพล

- จากสูตร

$$CI_{1-\alpha} \text{ of } \delta = d \pm \frac{z_{\alpha/2}}{\sqrt{n}}$$

$$CI_{1-.05} \text{ of } \delta = 0.5 \pm \frac{1.96}{\sqrt{100}}$$

$$CI_{.95} \text{ of } \delta = (0.304, 0.696)$$

จากข้อมูลทำให้มีความเชื่อมั่นระดับ .95 ว่าคอร์สอบรมภาษาอังกฤษ ทำให้คะแนนสอบเพิ่มขึ้นมีขนาดอิทธิพลระหว่าง 0.304 ถึง 0.696

กำลังทางสถิติ

- สมมติว่า คอร์สฝึกอบรม IQ หนึ่ง ทำให้คะแนน IQ เพิ่มขึ้น 6 แต้ม (μ = 106)
- สุ่มคนจำนวน 25 คน จากคอร์สดังกล่าว เพื่อดูว่าได้ค่าเฉลี่ย IQ เท่าไร
- ค่าเฉลี่ยของ IQ จากกลุ่มตัวอย่าง อาจจะได้ 102, 104, 106, 108, 110 ก็ได้
- ค่าเฉลี่ยของกลุ่มตัวอย่างที่ผ่านการฝึกอบรม จะไม่เท่ากับ 106 เสมอไป เนื่องจากเกิดความผิดพลาดจากการสุ่ม (Sampling Error)

กำลังทางสถิติ

- เนื่องจากค่าเฉลี่ยจากกลุ่มตัวอย่างที่สุ่มออกมานั้น แตกต่างกันไป
- ดังนั้น จึงเป็นไปได้ที่ ค่าเฉลี่ยจากกลุ่มตัวอย่างบางกลุ่มถึงระดับนัยสำคัญทางสถิติ และค่าเฉลี่ยจากกลุ่มตัวอย่างบางกลุ่มไม่ถึงระดับนัยสำคัญทางสถิติ
- โอกาสที่จะเจอค่าเฉลี่ยแตกต่างจากคนปกติถึงระดับนัยสำคัญทางสถิติ จากประชากรที่แตกต่างจริง (ในที่นี้ คือ ประชากรที่ผ่านการอบรม) จะเรียกว่า กำลังในการทดสอบทางสถิติ (Statistical Power)
- อาจเรียกสั้นๆ ได้ว่า กำลัง (Power)

กำลังทางสถิติ

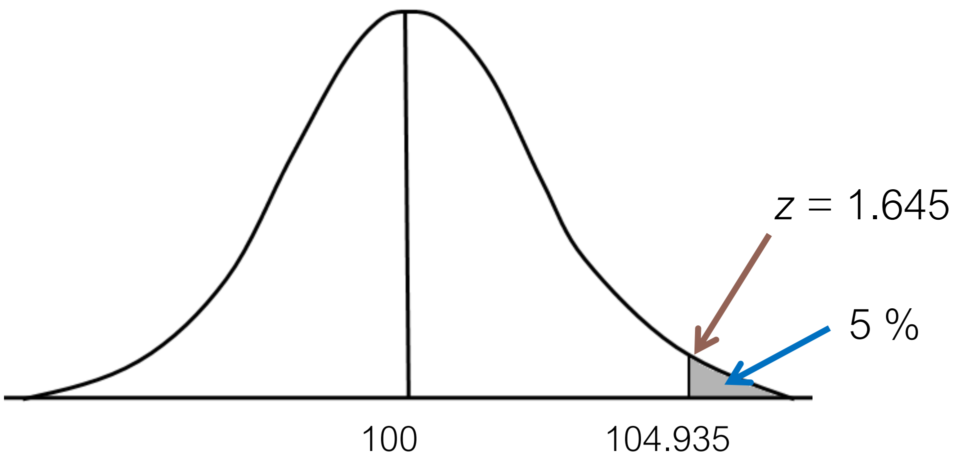
- ในการหากำลัง จะต้องทำ 2 ขั้นตอนกว้างๆ คือ
  1. หา Critical Value
  2. หาโอกาสที่จะสุ่มได้ค่าเฉลี่ยจากกลุ่มตัวอย่าง (ที่มาจากประชากรที่ Null Hypothesis ไม่เป็นจริง) สูงกว่า (หรือต่ำกว่า) Critical Value

กำลังทางสถิติ

- การหา Critical Value ทำเหมือนบทที่ผ่านมา
    1. สร้างการกระจายจาก Null Hypothesis
        - Population Distribution
        - Sampling Distribution
    2. นำค่า Alpha และข้อมูลการทดสอบทางเดียวหรือสองทาง มาหาค่า Critical Value
    3. นำค่า z มาแปลงเป็นคะแนนดิบ

กำลังทางสถิติ

- หาการกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง (Sampling Distribution of Means) ในกรณีที่สุ่มคนปกติ มาทีละ 25 คน
- จาก $\mu = 100$, $\sigma = 15$
- ได้ $\mu_M = 100$, $\sigma_M = 3$



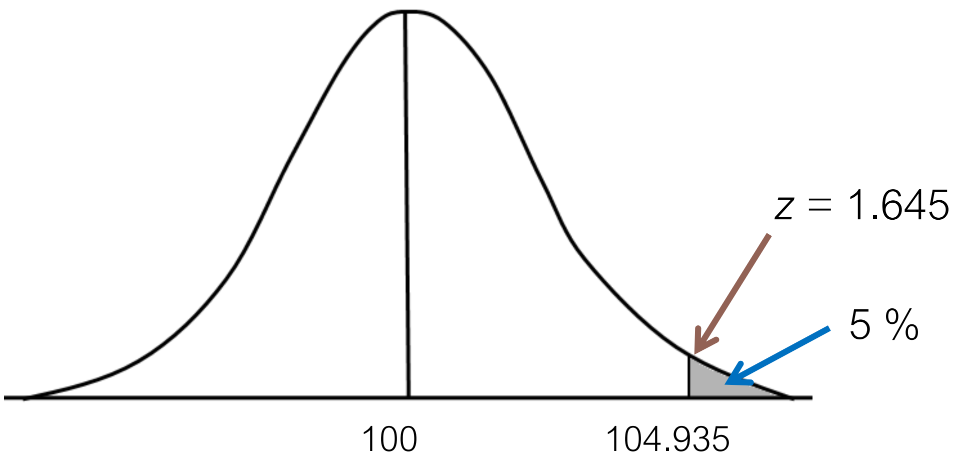
$$z = \frac{M - \mu_M}{\sigma_M}$$

$$1.645 = \frac{M - 100}{3}$$

$$M = 104.935$$

กำลังทางสถิติ

- Critical Value มีค่า IQ เท่ากับ 104.935 แสดงว่า
  - ถ้าค่าเฉลี่ย IQ ของผู้ผ่านการอบรม 25 คนสูงกว่า 104.935 จะแตกต่างอย่างมีนัยสำคัญ
  - ถ้าต่ำกว่า แสดงว่าไม่แตกต่างอย่างมีนัยสำคัญ



$$z = \frac{M - \mu_M}{\sigma_M}$$

$$1.645 = \frac{M - 100}{3}$$

$$M = 104.935$$

กำลังทางสถิติ

- หาโอกาสที่จะสุ่มได้ค่าเฉลี่ยจากกลุ่มตัวอย่างที่มีความแตกต่างจริง (Alternative Hypothesis เป็นจริง) แล้วได้ค่าเฉลี่ยสูงกว่า (หรือต่ำกว่า) Critical Value

1. สร้างการกระจายจาก Alternative Hypothesis
   - ❖ Population หรือ Sampling Distribution
   - ❖ กำหนดจากขนาดอิทธิพลที่กำหนด
2. นำ Critical Value มาแปลงเป็นค่า z ภายใต้การกระจายในข้อที่ 1
3. หาโอกาสที่จะได้ค่าเฉลี่ย สูงกว่า (หรือต่ำกว่า) ค่า *z* ดังกล่าว

กำลังทางสถิติ

- หาว่าโอกาสในการสุ่มกลุ่มตัวอย่าง 25 คน จากประชากรที่ผ่านการฝึกอบรมเพิ่มเชาวน์ปัญญา (μ = 106, σ = 15) แล้วเจอค่าเฉลี่ยสูงกว่าค่าวิกฤต (104.935) มีเท่าไร

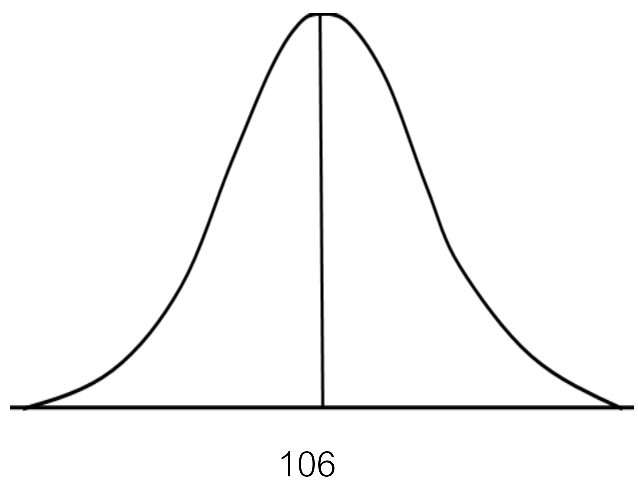
กำลังทางสถิติ

- การกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง

การกระจายของประชากร: $\mu = 106$, $\sigma = 15$

$n = 25$

การกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง: $\mu_M = 106$, $\sigma_M = 3$



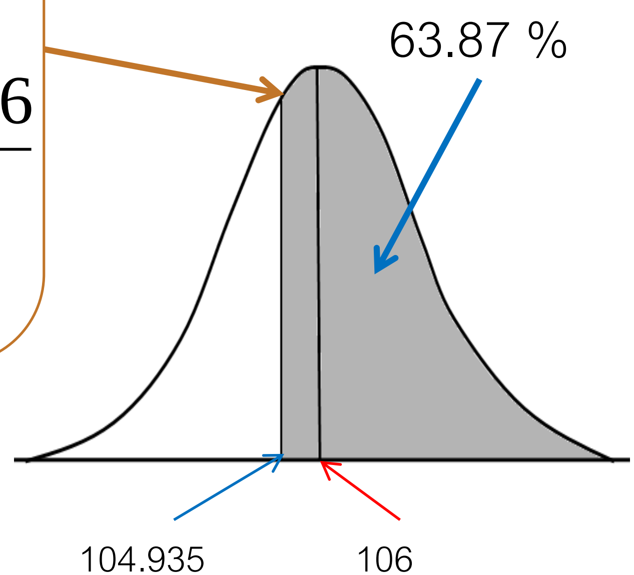
กำลังทางสถิติ

- หาโอกาสที่จะเจอกลุ่มตัวอย่างที่มีค่าเฉลี่ยถึงระดับนัยสำคัญ
- ก็คือ โอกาสที่จะเจอกลุ่มตัวอย่างที่มีค่าเฉลี่ยเกิน 104.935

$$z = \frac{M - \mu_M}{\sigma_M}$$

$$z = \frac{104.935 - 106}{3}$$

$$z = -0.355$$



โอกาสที่จะเจอกลุ่ม
ตัวอย่างที่มีค่าเฉลี่ยถึง
ระดับนัยสำคัญเท่ากับ
0.64 หรือ 64 %

กำลังทางสถิติ

สรุปอีกครั้งหนึ่ง

- โอกาสที่จะเจอกลุ่มตัวอย่างที่มีค่าเฉลี่ยถึงระดับนัยสำคัญ ***จากประชากรที่แตกต่างจริง*** จะเรียกว่า กำลังในการทดสอบทางสถิติ (Statistical Power) หรือเรียกสั้นๆ ว่ากำลัง (Power)
- ตรงกันข้ามกับ โอกาสที่จะเจอกลุ่มตัวอย่างที่มีค่าเฉลี่ยไม่ถึงระดับนัยสำคัญ ***จากประชากรที่แตกต่างจริง*** จะเรียกว่า โอกาสในการเจอความผิดพลาดแบบที่สอง (Type II error)

กำลังทางสถิติ

- ดังตาราง

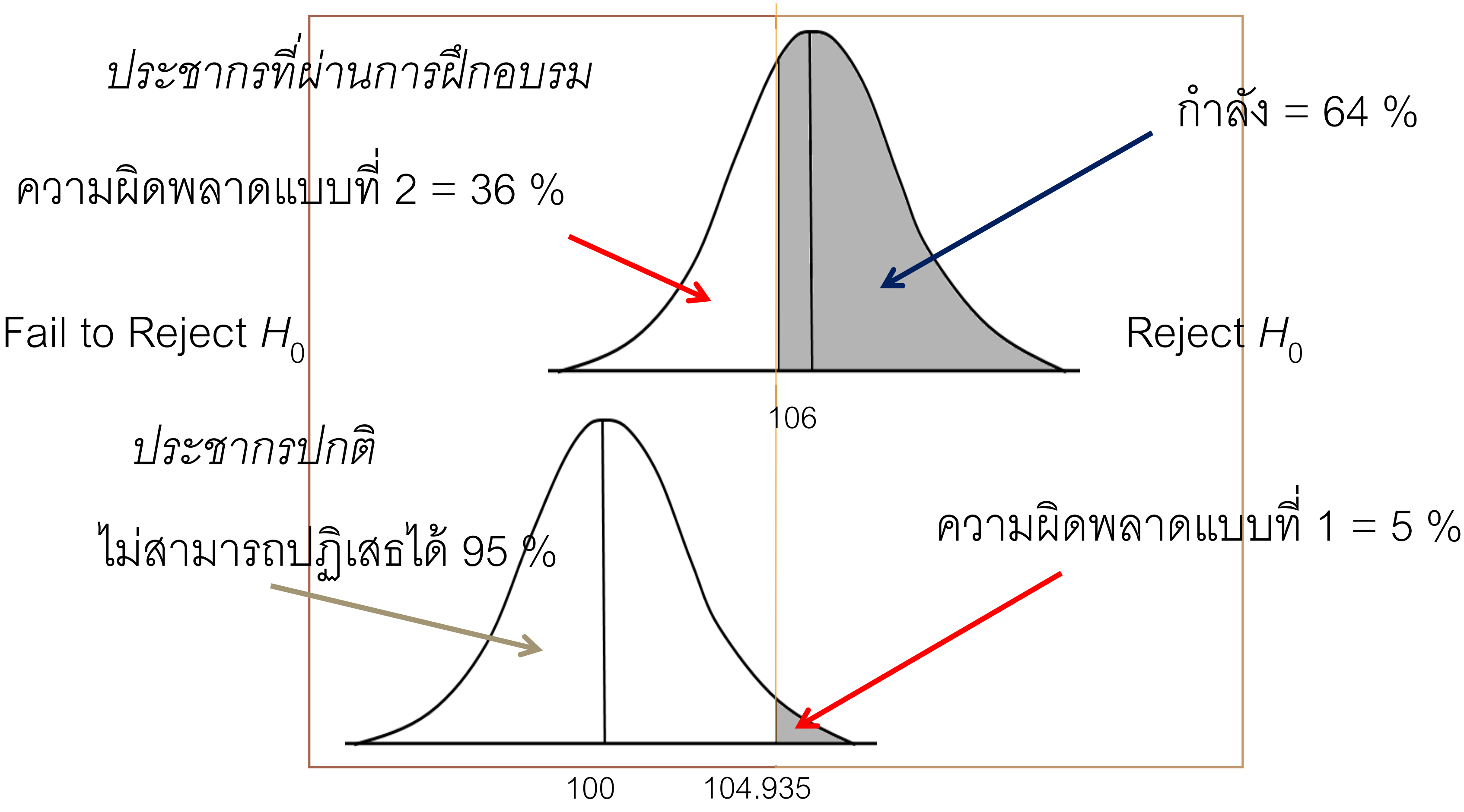
ความเป็นจริง

การตัดสินใจ

	H_0 ไม่เป็นจริง	H_0 เป็นจริง
Reject H_0	กำลัง (Power)	Type I error
Fail to Reject H_0	Type II error	???

กำลังทางสถิติ

- จากตัวอย่างที่ผ่านมา ลองดูภาพลักษณะการตัดสินสมมติฐาน

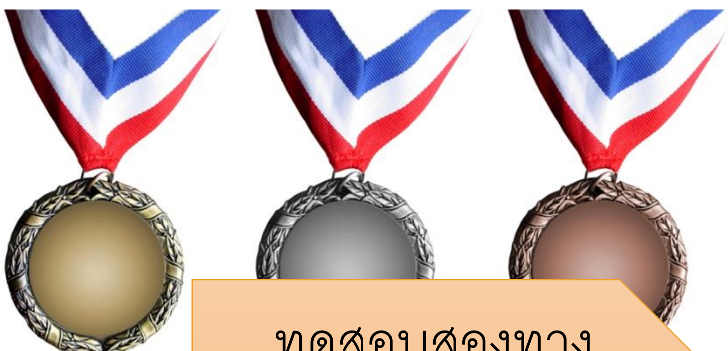


กำลังทางสถิติ

เรารู้อยู่แล้ว ว่าประชากรสองกลุ่มแตกต่างกันจริง
เพียงแต่ เราอยากรู้ว่าเมื่อทดสอบสมมติฐานแล้วจะเป็นอย่างไร

- ตัวอย่าง

นักกีฬาที่ได้เหรียญเงินจะมีความสุขน้อยกว่านักกีฬาที่แข่งขันทั่วไป



ทดสอบสองทาง
α = .05

นักกีฬาทั่วไป

μ = 50
σ = 10

นักกีฬาที่ได้เหรียญเงิน

μ = 40
σ = 10

ถ้าสุ่มนักกีฬาเหรียญเงิน มาทีละ 9 คน จะมีโอกาสที่จะสุ่มเจอกลุ่มตัวอย่างที่มีค่าเฉลี่ยถึงระดับนัยสำคัญ (กำลัง) เท่าไร

กำลังทางสถิติ

- ตัวอย่าง

นักกีฬาที่ได้เหรียญเงินจะมีความสุข
น้อยกว่านักกีฬาที่แข่งขันทั่วไป

นักกีฬาทั่วไป

นักกีฬาที่ได้เหรียญเงิน

ในการทดสอบ

สมมติฐานวิจัย

นักกีฬาเหรียญเงินมีความสุข
แตกต่างจากนักกีฬาทั่วไป

H_0: $\mu = 50$

H_1: $\mu \neq 50$

กำลังทางสถิติ

- ตัวอย่าง

นักกีฬาทั่วไป

นักกีฬาที่ได้เหรียญเงิน

นักกีฬาที่ได้เหรียญเงินจะมีความสุขน้อยกว่านักกีฬาที่แข่งขันทั่วไป

ความผิดพลาดแบบที่ 1	สุ่มประชากรนักกีฬาทั่วไป แต่ตัดสินว่าความสุขน้อยกว่านักกีฬาทั่วไป
ความผิดพลาดแบบที่ 2	สุ่มประชากรนักกีฬาที่ได้เหรียญเงิน แต่ตัดสินว่าไม่สามารถสรุปได้ว่านักกีฬาเหล่านี้ มีความสุขน้อยกว่านักกีฬาทั่วไปหรือไม่
กำลัง	สุ่มประชากรนักกีฬาที่ได้เหรียญเงิน แล้วตัดสินว่านักกีฬาเหล่านี้มีความสุขน้อยกว่านักกีฬาทั่วไปจริง

กำลังทางสถิติ

- ตัวอย่าง

หาค่าวิกฤตจากการเปรียบเทียบจากนักกีฬาทั่วไป

นักกีฬาที่ได้เหรียญเงินจะมีความสุข
น้อยกว่านักกีฬาที่แข่งขันทั่วไป

การกระจายประชากร

$\mu = 50; \sigma = 10$

สร้างการกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง

การกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง

$\mu_M = 50; \sigma_M = 3.33$

กำลังทางสถิติ

- ตัวอย่าง

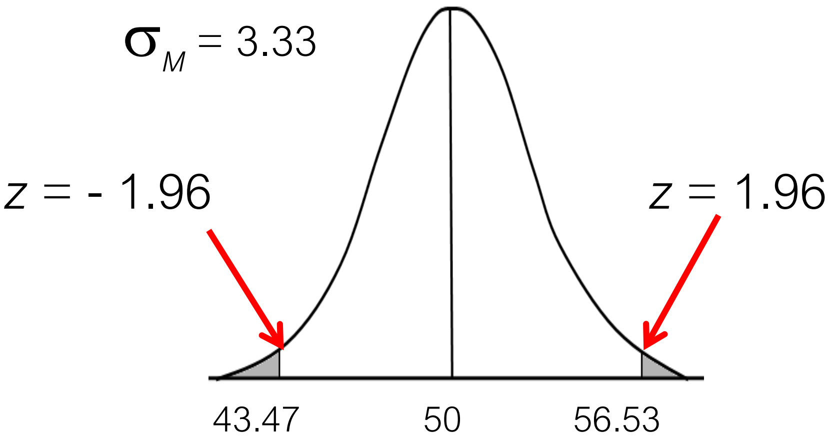
นักกีฬาที่ได้เหรียญเงินจะมีความสุข
น้อยกว่านักกีฬาที่แข่งขันทั่วไป

หาค่าวิกฤตจากการเปรียบเทียบจากนักกีฬาทั่วไป

α = .05; สองทาง

$\mu_M = 50$

$\sigma_M = 3.33$



กำลังทางสถิติ

- ตัวอย่าง

นักกีฬาที่ได้เหรียญเงินจะมีความสุข
น้อยกว่านักกีฬาที่แข่งขันทั่วไป

สร้างการกระจายค่าเฉลี่ยของกลุ่มตัวอย่างจากการสุ่มนักกีฬาที่ได้เหรียญเงิน ทีละ 9 คน

การกระจายประชากร

$\mu = 40; \sigma = 10$

สร้างการกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง

การกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง

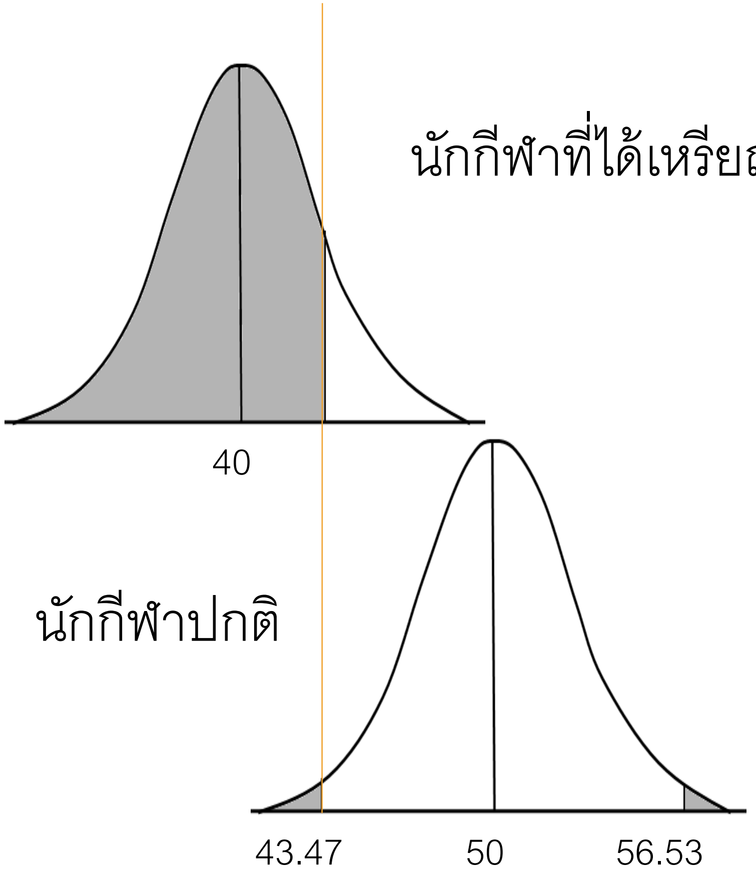
$\mu_M = 40; \sigma_M = 3.33$

กำลังทางสถิติ

- ตัวอย่าง

นักกีฬาที่ได้เหรียญเงินจะมีความสุข
น้อยกว่านักกีฬาที่แข่งขันทั่วไป

เปรียบเทียบการกระจายสองกลุ่ม

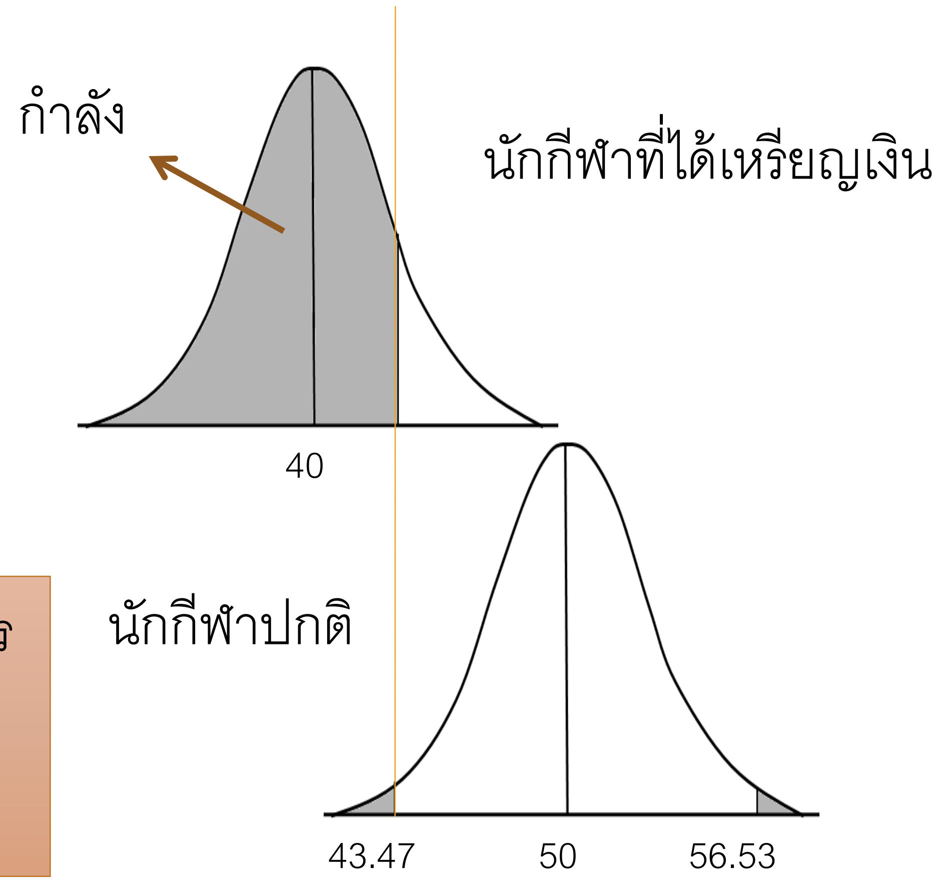


กำลังทางสถิติ

- ตัวอย่าง

หากำลังในการทดสอบสมมติฐานทางสถิติ

นักกีฬาที่ได้เหรียญเงินจะมีความสุข
น้อยกว่านักกีฬาที่แข่งขันทั่วไป



1. แปลงค่าจุดวิกฤต เป็นค่า z ในการกระจายของนักกีฬาเหรียญเงิน
2. หาพื้นที่แรงเงา

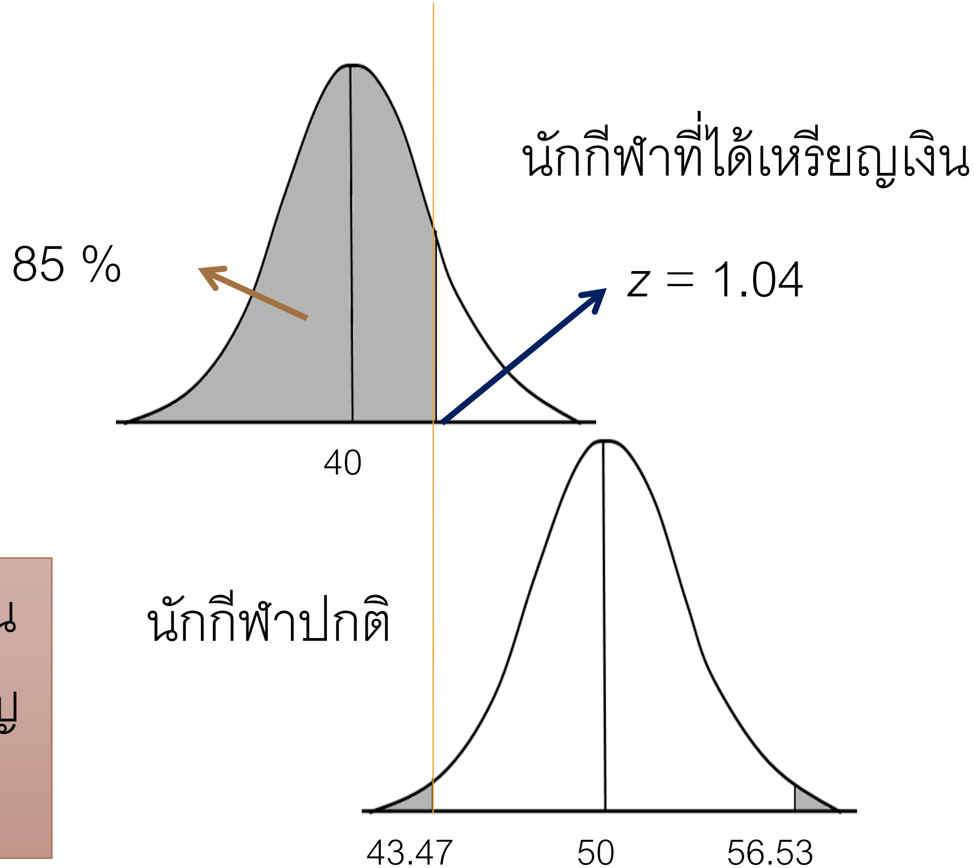
กำลังทางสถิติ

- ตัวอย่าง

หากำลังในการทดสอบสมมติฐานทางสถิติ

นักกีฬาที่ได้เหรียญเงินจะมีความสุข
น้อยกว่านักกีฬาที่แข่งขันทั่วไป

โอกาสที่สุ่มนักกีฬาเหรียญเงินจำนวน
9 คนแล้วจะแตกต่างอย่างมีนัยสำคัญ
เท่ากับ .85 หรือ 85 %



กำลังทางสถิติ

- สรุปวิธีการหากำลัง
  1. สร้างการกระจายค่าเฉลี่ยของกลุ่มอย่างจากการสุ่มจาก H_0
  2. หาค่าวิกฤตในรูปของคะแนนดิบ (Raw Score) จากการทดสอบตามระดับนัยสำคัญ และจำนวนทิศทางที่ต้องการ
  3. สร้างการกระจายค่าเฉลี่ยของกลุ่มตัวอย่างจากการสุ่มประชากรที่ใช้ในการหากำลัง
  4. นำการการกระจายค่าเฉลี่ยของกลุ่มตัวอย่างทั้งสองมาเปรียบเทียบกัน
  5. แปลงค่าวิกฤตในรูปคะแนนดิบให้เป็นค่ามาตรฐานในการกระจายในข้อที่ 3
  6. หากำลังในการทดสอบสมมติฐาน

กำลังทางสถิติ

- ตัวอย่าง

ผู้หญิงที่แต่งหน้าจะดูสวยขึ้น

ทดสอบทิศทางเดียว

α = .05

ผู้หญิงไม่แต่งหน้า

$\mu = 5$

$\sigma = 2$

ผู้หญิงแต่งหน้า

$\mu = 6$

$\sigma = 2$

ถ้าสุ่มผู้หญิงที่แต่งหน้า มาทีละ 4 คน จะมีโอกาสที่จะสุ่มเจอกลุ่มตัวอย่างที่มีค่าเฉลี่ยถึงระดับนัยสำคัญ (กำลัง) เท่าไร

กำลังทางสถิติ

- ตัวอย่าง

ผู้หญิงที่แต่งหน้าจะดูสวยขึ้น

ผู้หญิงไม่แต่งหน้า

ผู้หญิงแต่งหน้า

ในการทดสอบ

สมมติฐานวิจัย

ผู้หญิงที่แต่งหน้าสวยกว่าผู้หญิงที่ไม่แต่ง

$H_0: \mu \leq 5$

$H_1: \mu > 5$

กำลังทางสถิติ

- ตัวอย่าง

ผู้หญิงที่แต่งหน้าจะดูสวยขึ้น

หาการกระจายค่าเฉลี่ยของกลุ่มตัวอย่างจากผู้หญิงไม่แต่งหน้า

การกระจายประชากร

$\mu = 5; \sigma = 2$

สร้างการกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง

การกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง

$\mu_M = 5; \sigma_M = 1$

กำลังทางสถิติ

- ตัวอย่าง

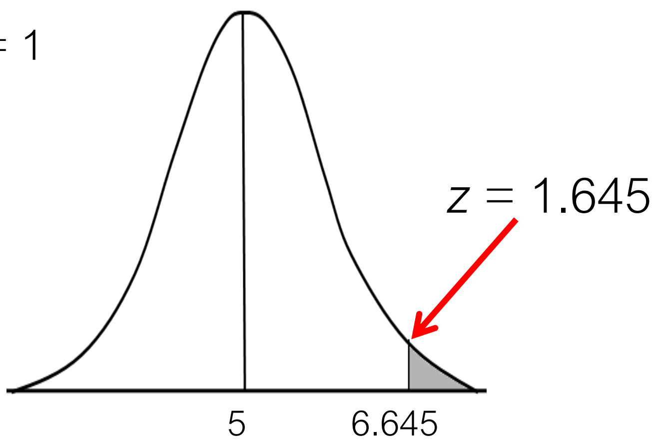
หาค่าวิกฤตจากการเปรียบเทียบกับผู้หญิงไม่แต่งหน้า

α = .05; ทางเดียว

ผู้หญิงที่แต่งหน้าจะดูสวยขึ้น

$\mu_M = 5$

$\sigma_M = 1$



กำลังทางสถิติ

- ตัวอย่าง

สร้างการกระจายค่าเฉลี่ยของกลุ่มตัวอย่างจากการสุ่มผู้หญิงแต่งหน้า ทีละ 4 คน

ผู้หญิงที่แต่งหน้าจะดูสวยขึ้น

การกระจายประชากร

$\mu = 6; \sigma = 2$

↓

สร้างการกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง

↓

การกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง

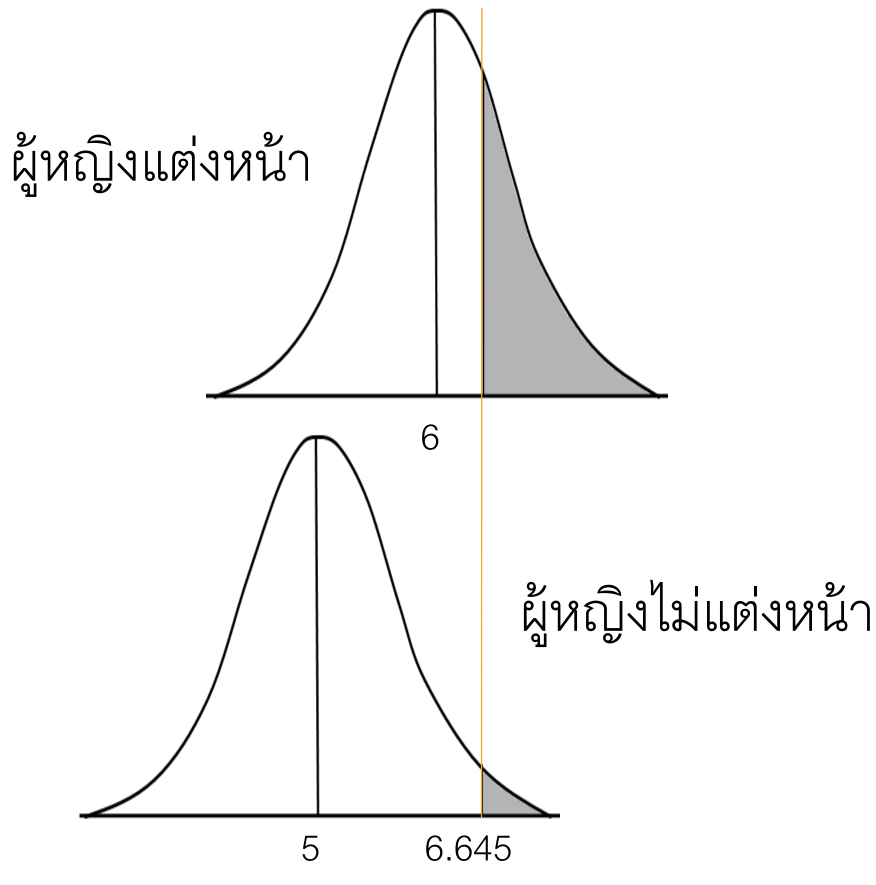
$\mu_M = 6; \sigma_M = 1$

กำลังทางสถิติ

- ตัวอย่าง

ผู้หญิงที่แต่งหน้าจะดูสวยขึ้น

เปรียบเทียบการกระจายสองกลุ่ม

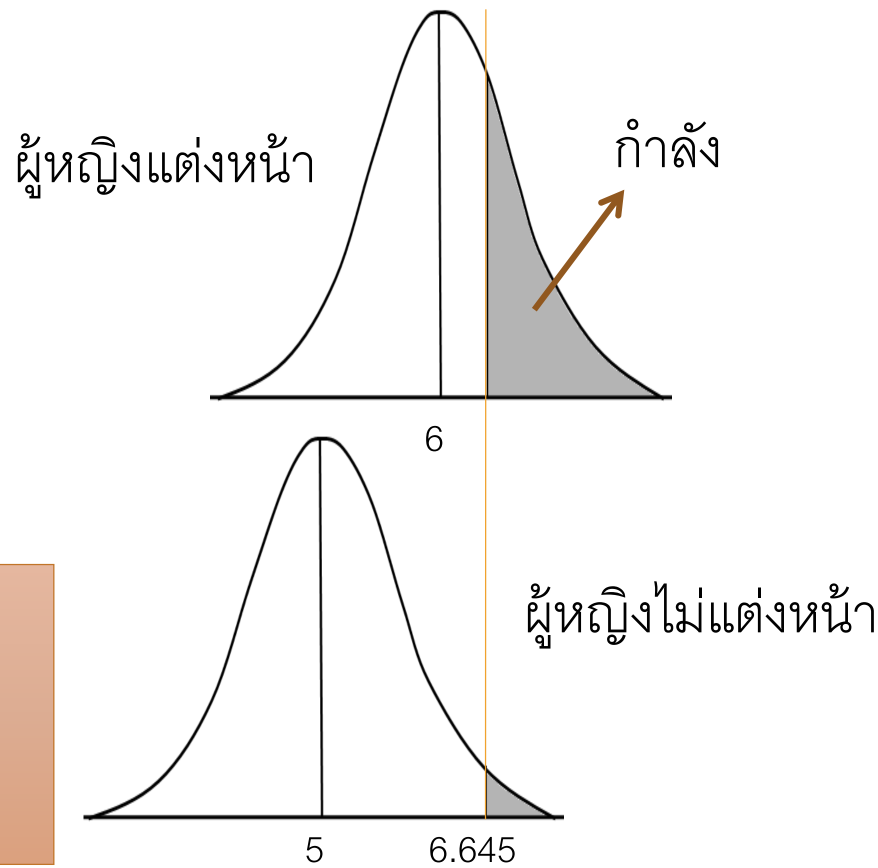


กำลังทางสถิติ

- ตัวอย่าง

หากำลังในการทดสอบสมมติฐานทางสถิติ

ผู้หญิงที่แต่งหน้าจะดูสวยขึ้น



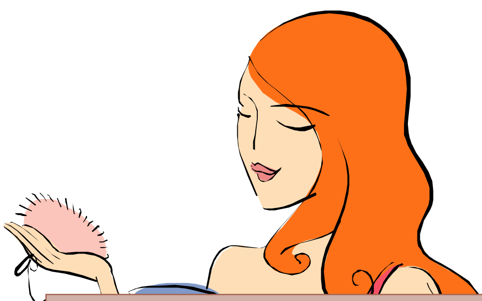
1. แปลงค่าจุดวิกฤต เป็นค่า z ใน การกระจายของผู้หญิงแต่งหน้า
2. หาพื้นที่แรงเงา

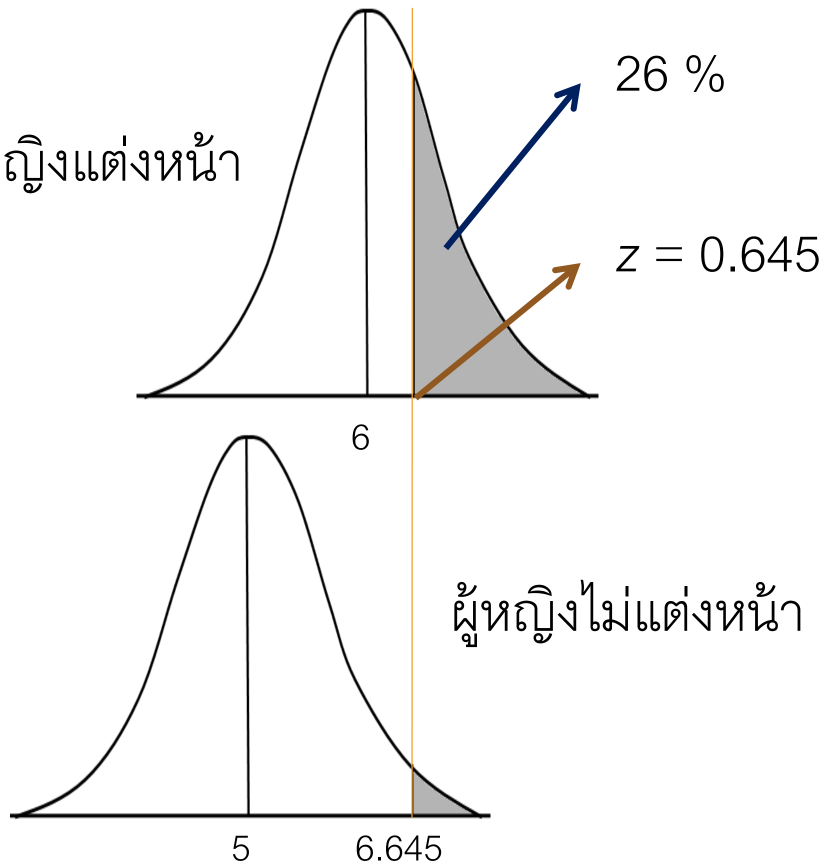
กำลังทางสถิติ

- ตัวอย่าง

หากำลังในการทดสอบสมมติฐานทางสถิติ

ผู้หญิงที่แต่งหน้าจะดูสวยขึ้น





โอกาสที่สุ่มผู้หญิงแต่งหน้าจำนวน 4 คนแล้วจะแตกต่างอย่างมีนัยสำคัญเท่ากับ .26 หรือ 26 %

ปัจจัยที่มีผลต่อกำลังทางสถิติ

- ในการวิจัย ต้องการให้การทดสอบสมมติฐานทางสถิติมีกำลังสูง
- ทำอย่างไรที่จะทำให้การทดสอบมีกำลังสูง
- ปัจจัยที่ทำให้เปลี่ยนกำลัง
  - ความแตกต่างระหว่างค่าเฉลี่ย
  - การกระจายของประชากร
  - จำนวนกลุ่มตัวอย่าง
  - ระดับนัยสำคัญ
  - จำนวนทิศทาง

ขนาดอิทธิพล

ปัจจัยที่มีผลต่อกำลังทางสถิติ

- ความแตกต่างระหว่างค่าเฉลี่ย

ทดสอบทิศทางเดียว; α = .05; n = 4

การกระจายประชากร

ผู้หญิงแต่งหน้า

$\mu = 6; \sigma = 2$

การกระจายประชากร

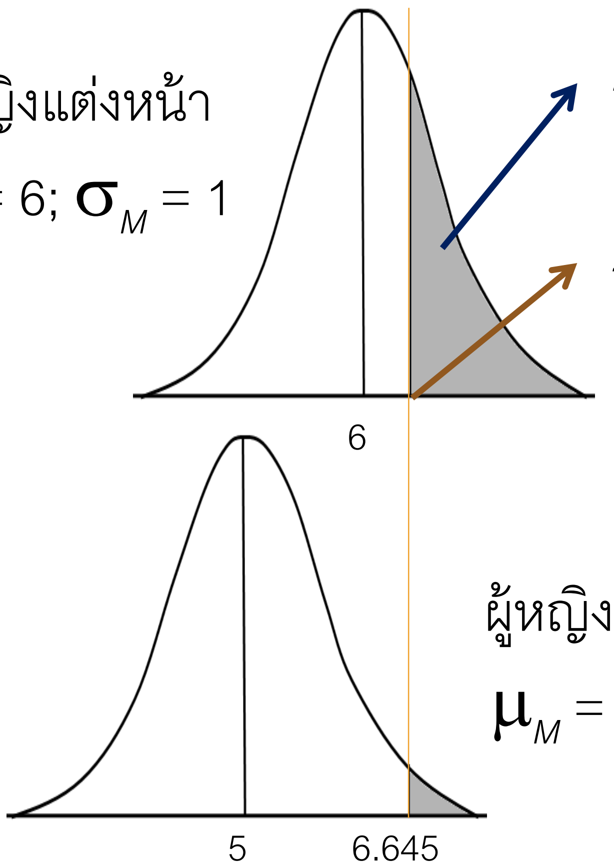
ผู้หญิงไม่แต่งหน้า

$\mu = 5; \sigma = 2$

แตกต่าง 1 แต้ม; $d = 0.5$

ผู้หญิงแต่งหน้า

$\mu_M = 6; \sigma_M = 1$



ผู้หญิงไม่แต่งหน้า

$\mu_M = 5; \sigma_M = 1$

ปัจจัยที่มีผลต่อกำลังทางสถิติ

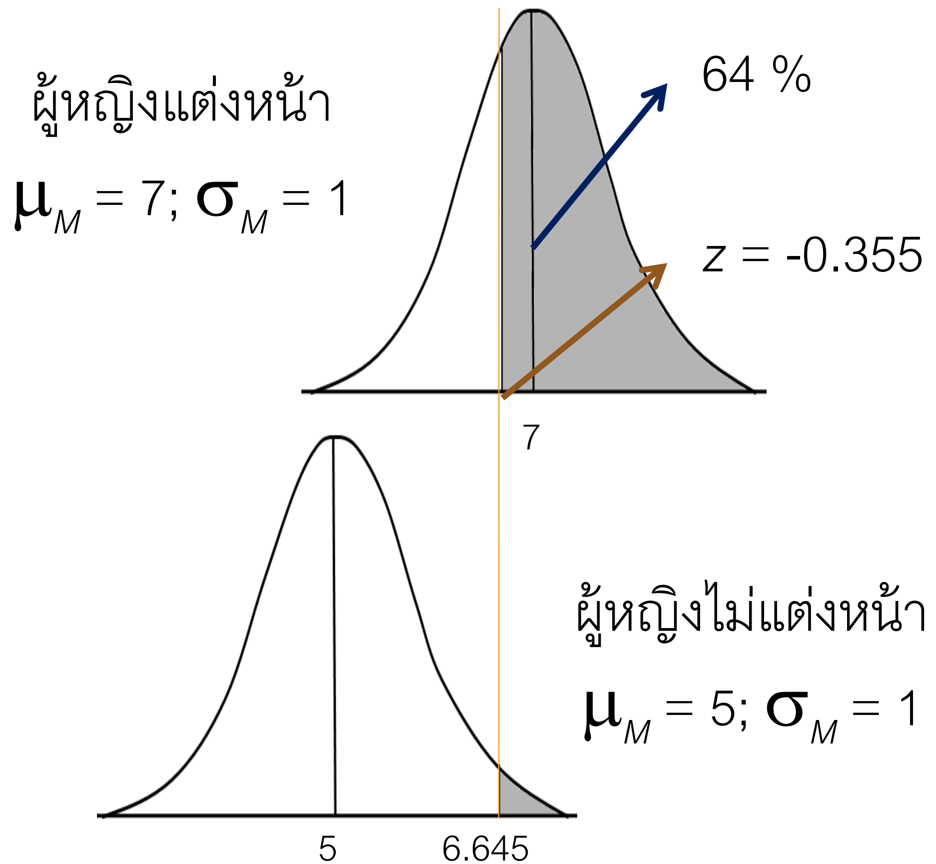
- ความแตกต่างระหว่างค่าเฉลี่ย

ทดสอบทิศทางเดียว; α = .05; n = 4

การกระจายประชากร
ผู้หญิงแต่งหน้า
$\mu = 7; \sigma = 2$

การกระจายประชากร
ผู้หญิงไม่แต่งหน้า
$\mu = 5; \sigma = 2$

แตกต่าง 2 แต้ม; $d = 1$



ปัจจัยที่มีผลต่อกำลังทางสถิติ

- ความแตกต่างของค่าเฉลี่ย

แตกต่าง 1 แต้ม

$d = 0.5$ → กำลัง = 26 %

แตกต่าง 2 แต้ม

$d = 1$ → กำลัง = 64 %

- ยิ่งความแตกต่างมากขึ้น หรือขนาดอิทธิพลเพิ่มขึ้น กำลังจะเพิ่มขึ้น
- ในการทดลอง จะต้องพยายามให้กลุ่มที่เปรียบเทียบแตกต่างมากที่สุด เพื่อให้มีกำลังในการทดสอบทางสถิติมากที่สุด เช่น แต่งหน้าฝีมือดีที่สุด

ปัจจัยที่มีผลต่อกำลังทางสถิติ

- การกระจายของประชากร

ทดสอบทิศทางเดียว; $\alpha = .05$; $n = 4$

การกระจายประชากร
ผู้หญิงแต่งหน้า
$\mu = 6$; $\sigma = 2$

การกระจายประชากร
ผู้หญิงไม่แต่งหน้า
$\mu = 5$; $\sigma = 2$

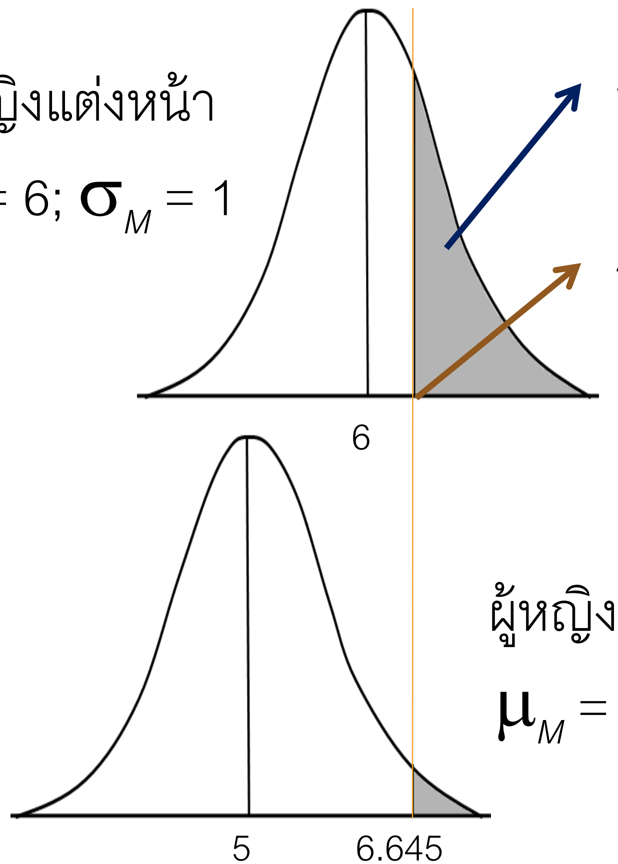
$\sigma = 2$; $d = 0.5$

ผู้หญิงแต่งหน้า

$\mu_M = 6$; $\sigma_M = 1$

26 %

$z = 0.645$



ผู้หญิงไม่แต่งหน้า

$\mu_M = 5$; $\sigma_M = 1$

ปัจจัยที่มีผลต่อกำลังทางสถิติ

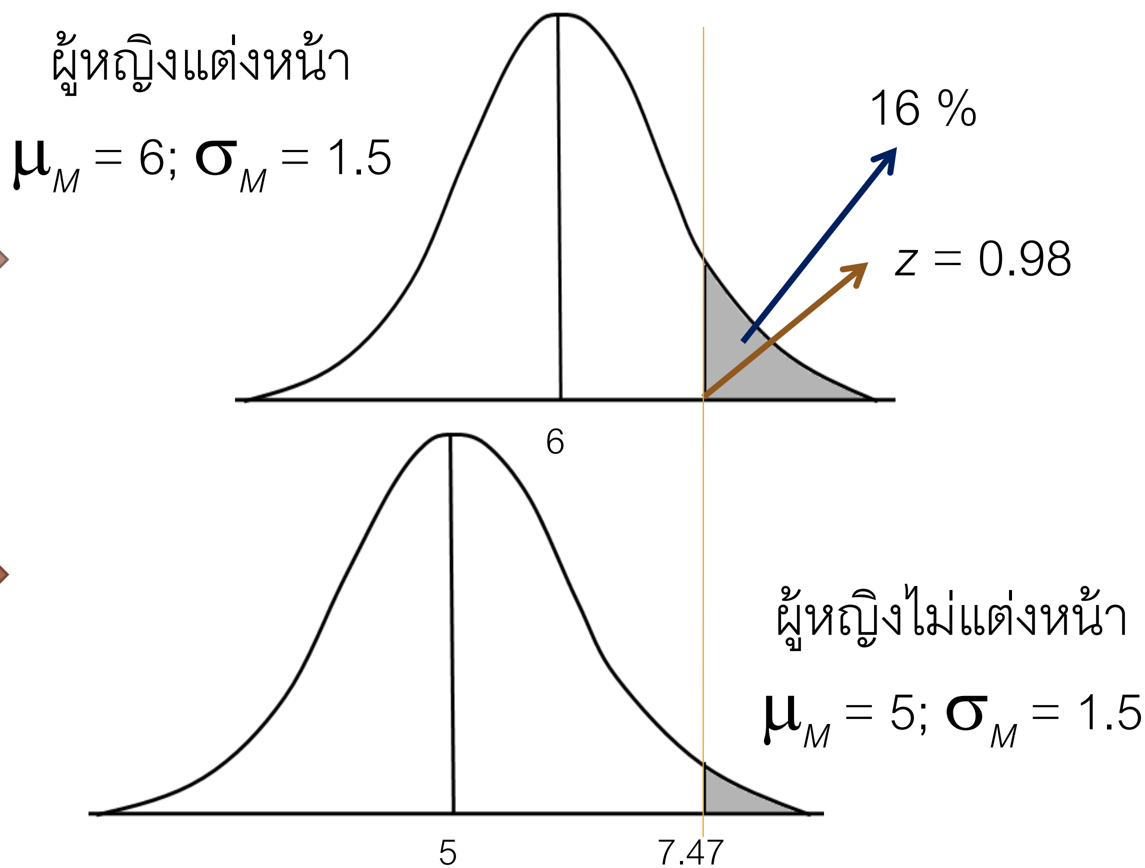
- การกระจายของประชากร

ทดสอบทิศทางเดียว; α = .05; n = 4

การกระจายประชากร
ผู้หญิงแต่งหน้า
μ = 6; σ = 3

การกระจายประชากร
ผู้หญิงไม่แต่งหน้า
μ = 5; σ = 3

σ = 3; d = 0.33



ปัจจัยที่มีผลต่อกำลังทางสถิติ

- การกระจายของประชากร

$\sigma = 2$, $d = 0.5$ → กำลัง = 26 %

$\sigma = 3$, $d = 0.33$ → กำลัง = 16 %

- ยิ่งการกระจายมากขึ้น หรือขนาดอิทธิพลลดลง กำลังจะลดลง
- ในการทดลอง จะต้องควบคุมให้มีตัวแปรแทรกซ้อนให้คงที่ เพราะตัวแปรเหล่านี้จะทำให้การกระจายเพิ่มขึ้นโดยไม่จำเป็น

ปัจจัยที่มีผลต่อกำลังทางสถิติ

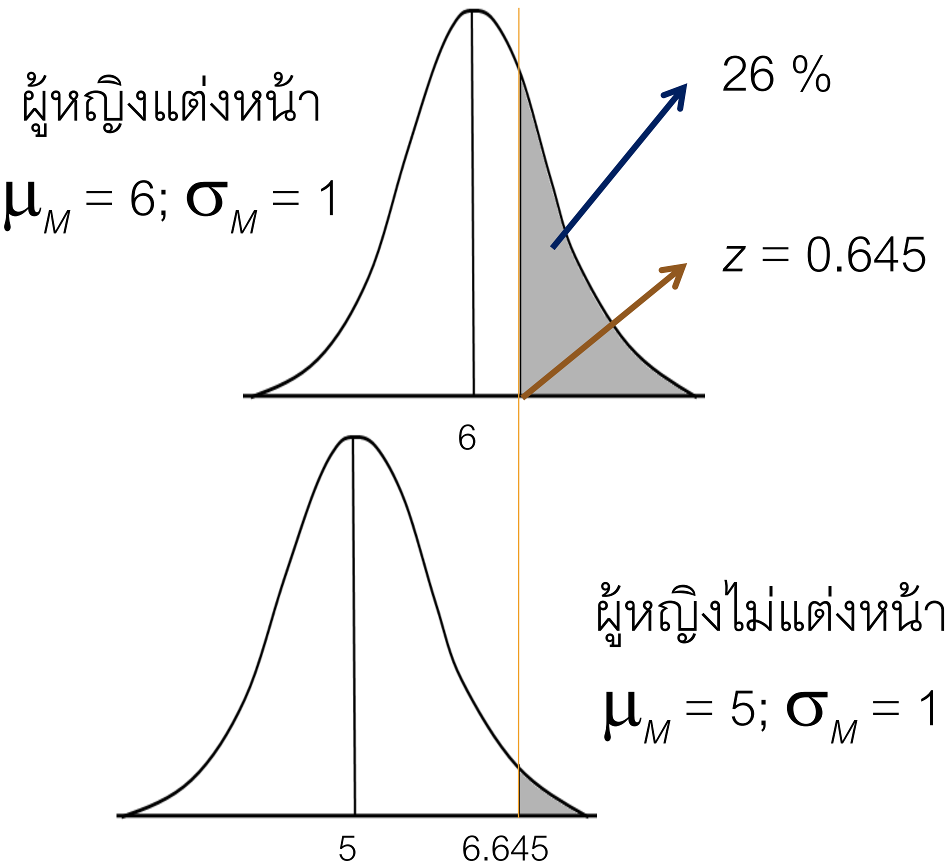
- จำนวนกลุ่มตัวอย่าง

ทดสอบทิศทางเดียว; α = .05; $n = 4$

การกระจายประชากร
ผู้หญิงแต่งหน้า
$\mu = 6; \sigma = 2$

การกระจายประชากร
ผู้หญิงไม่แต่งหน้า
$\mu = 5; \sigma = 2$

$n = 4$



ปัจจัยที่มีผลต่อกำลังทางสถิติ

- จำนวนกลุ่มตัวอย่าง

ทดสอบทิศทางเดียว; α = .05; *n* = 16

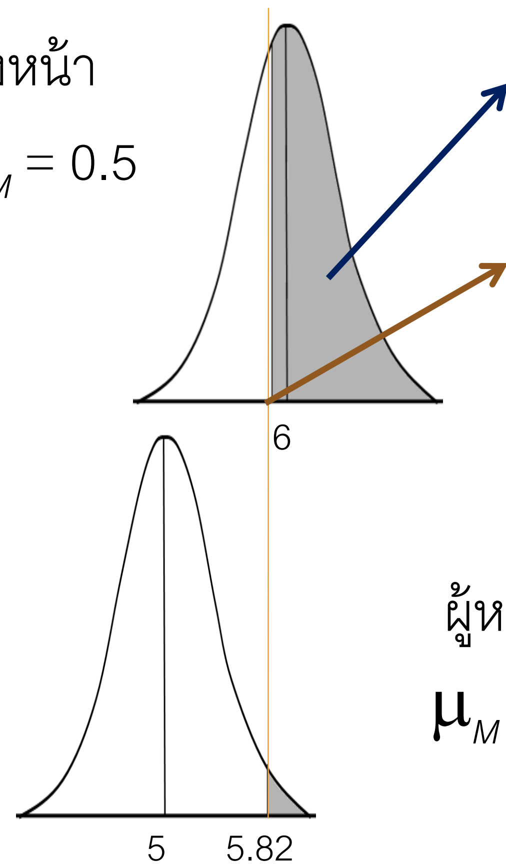
การกระจายประชากร
ผู้หญิงแต่งหน้า
$\mu = 6; \sigma = 2$

การกระจายประชากร
ผู้หญิงไม่แต่งหน้า
$\mu = 5; \sigma = 2$

n = 16

ผู้หญิงแต่งหน้า

$\mu_M = 6; \sigma_M = 0.5$



64 %

$z = -0.355$

ผู้หญิงไม่แต่งหน้า

$\mu_M = 5; \sigma_M = 0.5$

ปัจจัยที่มีผลต่อกำลังทางสถิติ

- จำนวนกลุ่มตัวอย่าง

$n = 4$ → กำลัง = 26 %

$n = 16$ → กำลัง = 64 %

- ยิ่งจำนวนกลุ่มตัวอย่างมากขึ้น กำลังจะมากขึ้น
- ในการทดลอง จะต้องเก็บข้อมูลจนกระทั่งแน่ใจว่ามีกำลังในการทดสอบสมมติฐานทางสถิติมากเพียงพอ

ปัจจัยที่มีผลต่อกำลังทางสถิติ

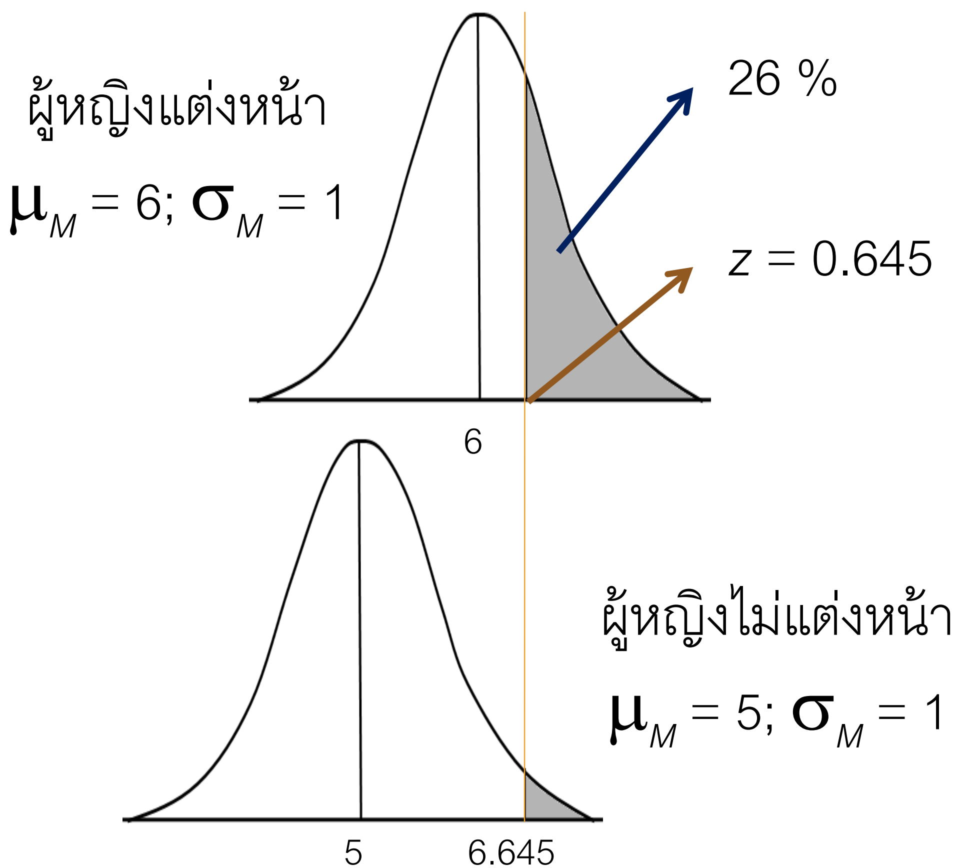
- จำนวนทิศทาง

ทดสอบ**ทิศทางเดียว**; α = .05; n = 4

การกระจายประชากร
ผู้หญิงแต่งหน้า
$\mu = 6; \sigma = 2$

การกระจายประชากร
ผู้หญิงไม่แต่งหน้า
$\mu = 5; \sigma = 2$

ทางเดียว



ปัจจัยที่มีผลต่อกำลังทางสถิติ

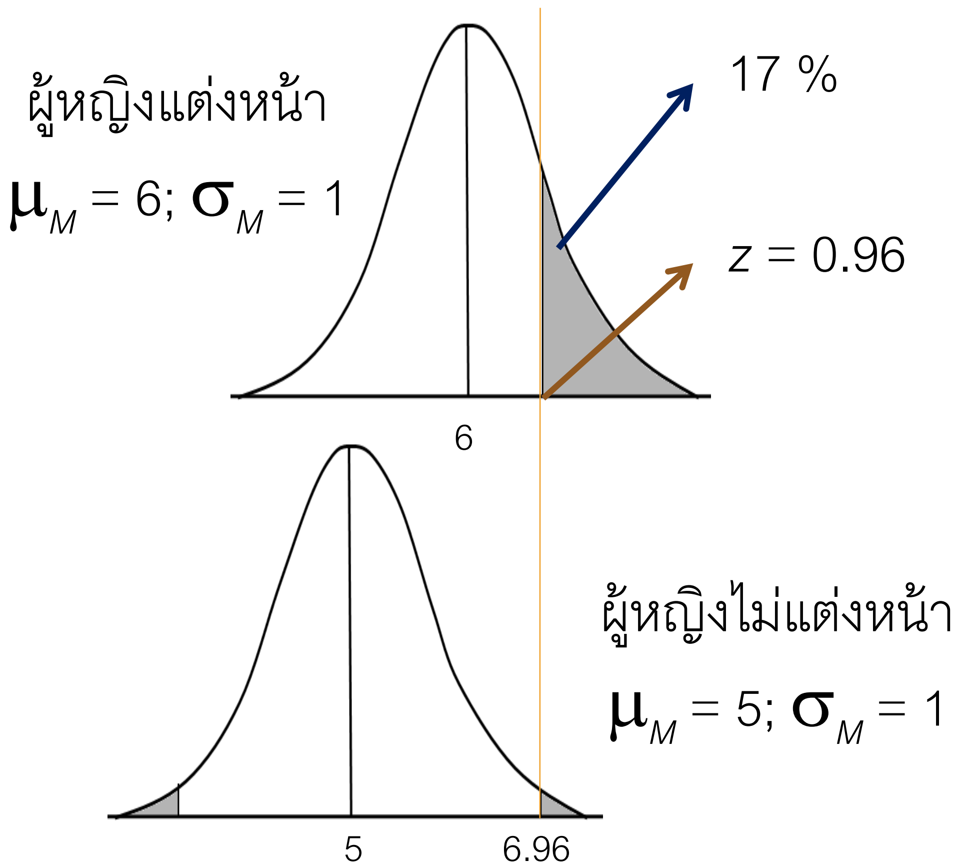
- จำนวนทิศทาง

ทดสอบ**สองทาง**; α = .05; n = 4

การกระจายประชากร
ผู้หญิงแต่งหน้า
μ = 6; σ = 2

การกระจายประชากร
ผู้หญิงไม่แต่งหน้า
μ = 5; σ = 2

สองทาง



ปัจจัยที่มีผลต่อกำลังทางสถิติ

- จำนวนทิศทาง

ทางเดียว → กำลัง = 26 %

สองทาง → กำลัง = 17 %

- การทดสอบสองทางจะมีกำลังน้อยกว่าการทดสอบทางเดียวเสมอ
- ถ้าสมมติฐานที่ใช้ทดสอบ สอดคล้องกับทิศทางของความแตกต่างของประชากร

ปัจจัยที่มีผลต่อกำลังทางสถิติ

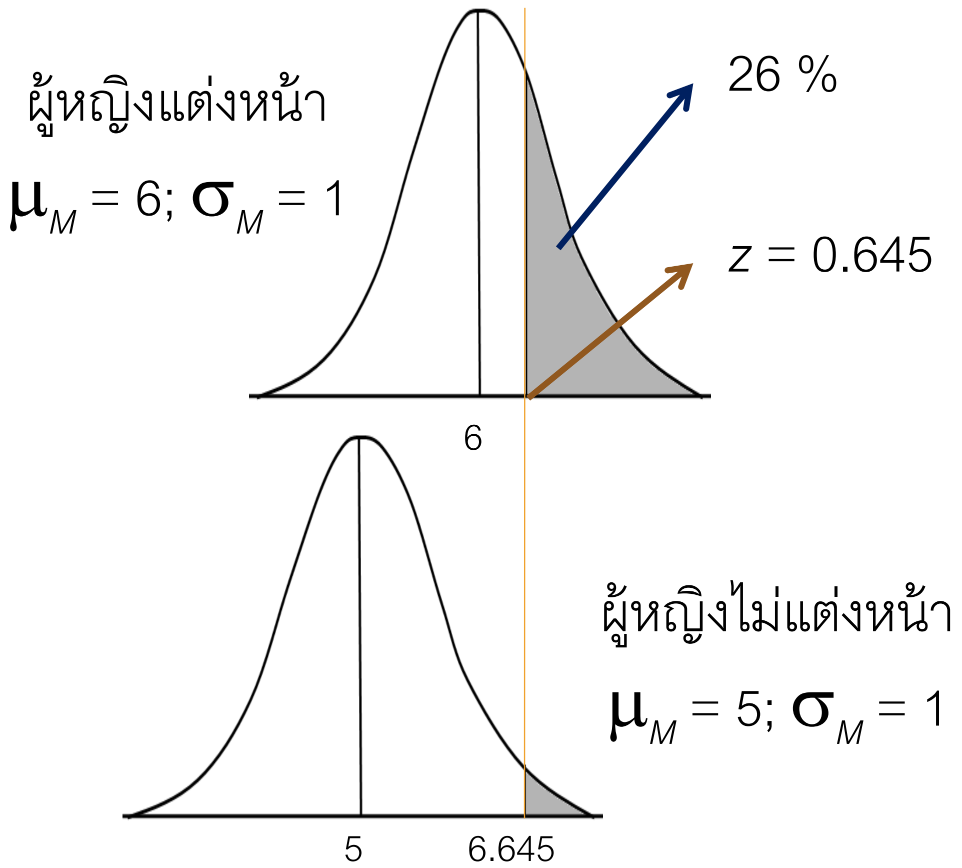
- ระดับนัยสำคัญ

ทดสอบทิศทางเดียว; α = .05; n = 4

การกระจายประชากร
ผู้หญิงแต่งหน้า
$\mu = 6; \sigma = 2$

การกระจายประชากร
ผู้หญิงไม่แต่งหน้า
$\mu = 5; \sigma = 2$

$\alpha = .05$



ปัจจัยที่มีผลต่อกำลังทางสถิติ

- ระดับนัยสำคัญ

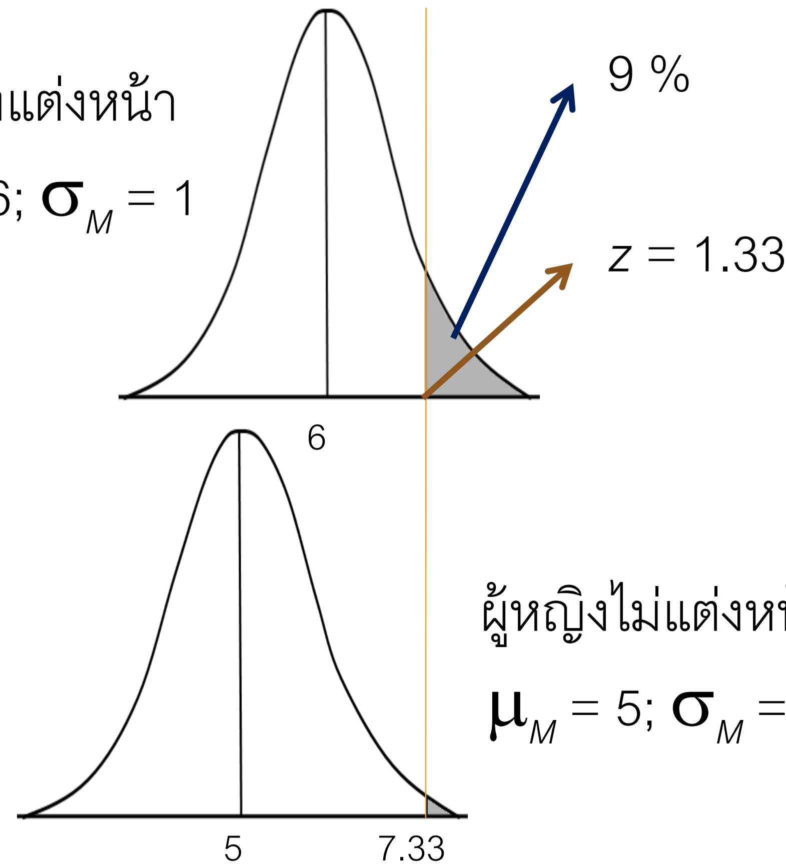
ทดสอบทิศทางเดียว; $\alpha = .01$; $n = 4$

การกระจายประชากร
ผู้หญิงแต่งหน้า
$\mu = 6; \sigma = 2$

การกระจายประชากร
ผู้หญิงไม่แต่งหน้า
$\mu = 5; \sigma = 2$

$\alpha = .01$

ผู้หญิงแต่งหน้า
$\mu_M = 6; \sigma_M = 1$



ผู้หญิงไม่แต่งหน้า
$\mu_M = 5; \sigma_M = 1$

ปัจจัยที่มีผลต่อกำลังทางสถิติ

- จำนวนทิศทาง

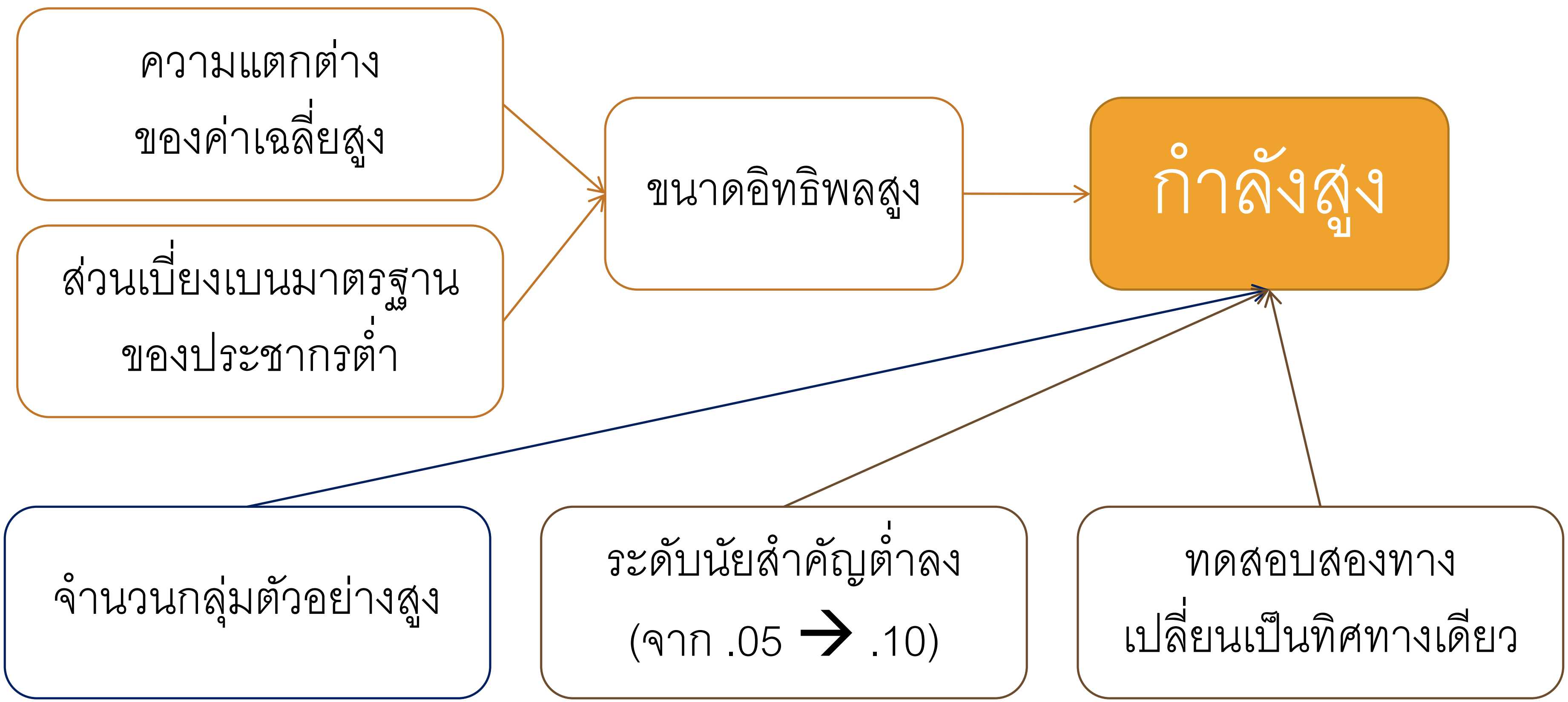
α = .05 → กำลัง = 26 %

α = .01 → กำลัง = 9 %

- ยิ่งตั้งระดับนัยสำคัญให้สูงเท่าไร (จาก .05 → .01) จะทำให้กำลังลดลง

ปัจจัยที่มีผลต่อกำลังทางสถิติ

- สรุปปัจจัยที่ทำให้เปลี่ยนกำลัง



การกำหนดขนาดกลุ่มตัวอย่าง

- ควรเก็บข้อมูลให้มีขนาดมากเพียงพอที่จะทำให้โอกาสในการปฏิเสธ Null Hypothesis สูง
- กล่าวคือ มีกำลังในการปฏิเสธสมมติฐานสูง เช่น กำลังมีสูงเท่ากับ .80 หรือ .90
- จากความรู้เรื่องการหากำลัง จะทำให้หาขนาดกลุ่มตัวอย่างที่สามารถทำให้ได้กำลังขนาดที่ต้องการได้ เมื่อทราบความแตกต่างของค่าเฉลี่ย หรือขนาดอิทธิพล ว่ามีขนาดเท่าไร

การกำหนดขนาดกลุ่มตัวอย่าง

- เช่น คอร์สฝึกอบรมโดยปกติ จะทำให้ค่า IQ เพิ่มขึ้น 10 แต้ม ควรจะเก็บข้อมูลเท่าไร ให้มีกำลังเท่ากับ .80
- ประชากรปกติมีคะแนน IQ: μ = 100, σ = 15
- ใช้การทดสอบสองทาง และตั้ง α = .05

การกำหนดขนาดกลุ่มตัวอย่าง

- การหากำลังจะต้องสร้างการกระจายค่าเฉลี่ยของกลุ่มตัวอย่างจากประชากรปกติ และประชากรที่ผ่านการฝึกอบรม

ประชากรปกติ	ประชากรที่ผ่านการฝึกอบรม
$\mu = 100; \sigma = 15$	$\mu = 110; \sigma = 15$
↓	↓
สร้างการกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง	สร้างการกระจายค่าเฉลี่ยของกลุ่มตัวอย่าง
↓	↓
$\mu_M = 100; \sigma_M = 15/\sqrt{n}$	$\mu_M = 110; \sigma_M = 15/\sqrt{n}$

การกำหนดขนาดกลุ่มตัวอย่าง

- หาจุดวิกฤตที่ทำให้ α = .05 (สองทาง) และกำลัง = .80

ประชากรที่ผ่านการฝึกอบรม

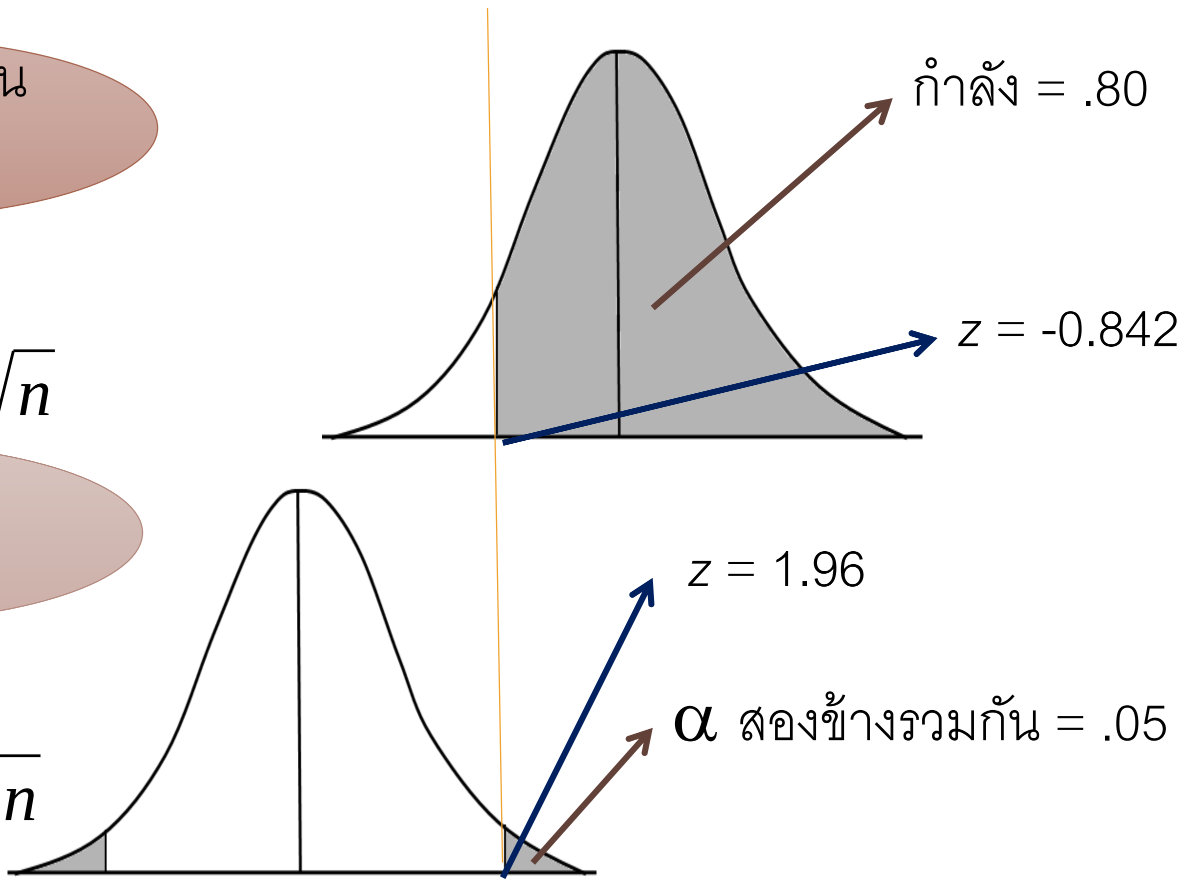
$$\mu_M = 110$$

$$\sigma_M = 15/\sqrt{n}$$

ประชากรปกติ

$$\mu_M = 100$$

$$\sigma_M = 15/\sqrt{n}$$

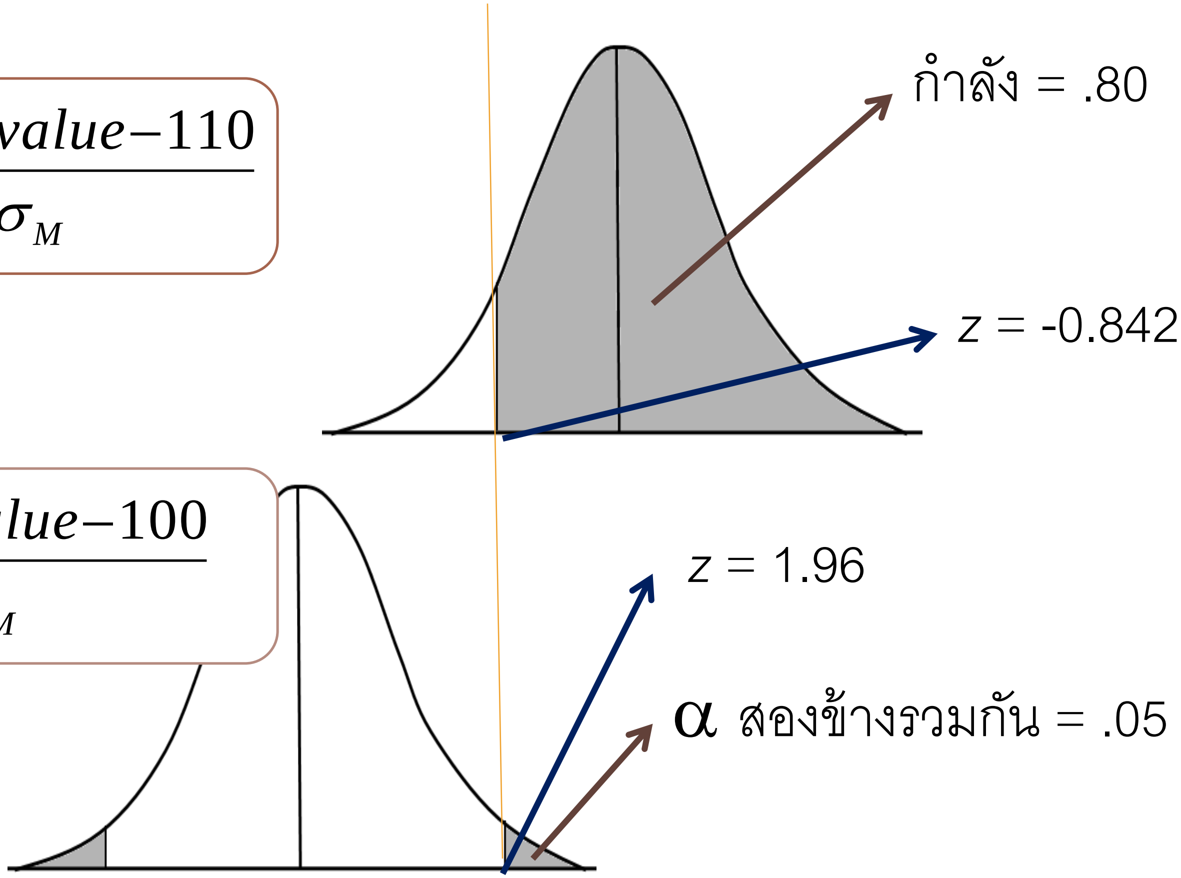


การกำหนดขนาดกลุ่มตัวอย่าง

- หาจุดวิกฤตที่ทำให้ α = .05 (สองทาง) และกำลัง = .80

$$-0.842 = \frac{critical\,value - 110}{\sigma_M}$$

$$1.96 = \frac{critical\,value - 100}{\sigma_M}$$



การกำหนดขนาดกลุ่มตัวอย่าง

- หาจุดวิกฤตที่ทำให้ α = .05 (สองทาง) และกำลัง = .80

ประชากรปกติ

ประชากรที่ผ่านการฝึกอบรม

$$1.96 = \frac{critical\,value - 100}{\sigma_M}$$

$$-0.842 = \frac{critical\,value - 110}{\sigma_M}$$

$$critical\,value = 1.96\sigma_M + 100$$

$$critical\,value = -0.842\sigma_M + 110$$

$$1.96\sigma_M + 100 = -0.842\sigma_M + 110$$

$$\boxed{\sigma_M = 3.567}$$

การกำหนดขนาดกลุ่มตัวอย่าง

- หาจุดวิกฤตที่ทำให้ α = .05 (สองทาง) และกำลัง = .80

$$\sigma_M = 3.567$$

$$\sigma_M = \sigma / \sqrt{n}$$

$$n = \left(\frac{\sigma}{\sigma_M} \right)^2 = \left(\frac{15}{3.567} \right)^2 \approx 18$$

จะต้องเก็บกลุ่มตัวอย่างอย่างน้อย 18 คน ถึงจะทำให้โอกาสสุ่มคนจากประชากรที่ฝึกอบรมมีค่าเฉลี่ยถึงระดับนัยสำคัญเกิน 80 %

การกำหนดขนาดกลุ่มตัวอย่าง

- ลองหาสูตร

ประชากรปกติ

ประชากรที่ผ่านการฝึกอบรม

$$z_{H0} = \frac{critical\,value - \mu_{H0}}{\sigma_M}$$

$$z_{H1} = \frac{critical\,value - \mu_{H1}}{\sigma_M}$$

$$critical\,value = z_{H0}\sigma_M + \mu_{H0}$$

$$critical\,value = z_{H1}\sigma_M + \mu_{H1}$$

$$z_{H0}\sigma_M + \mu_{H0} = z_{H1}\sigma_M + \mu_{H1}$$

$$\sigma_M = \frac{\mu_{H0} - \mu_{H1}}{z_{H1} - z_{H0}}$$

$$n = \left(\frac{\sigma(z_{H1} - z_{H0})}{\mu_{H0} - \mu_{H1}}\right)^2$$

การกำหนดขนาดกลุ่มตัวอย่าง

- รู้ความแตกต่างระหว่างค่าเฉลี่ยได้อย่างไร
  - งานวิจัยเก่า ว่าเคยได้ความแตกต่างเท่าไร หรือมีขนาดอิทธิพลเท่าไร
  - ตั้งเกณฑ์ที่เรียกว่าสำคัญ (Practical Significance)
    - เช่น การบำบัดนี้จะได้ผลคุ้มค่าเงิน จะต้องทำให้คะแนน IQ เพิ่มขึ้นอย่างน้อย 10 คะแนน
    - การบำบัดความซึมเศร้าจะได้ผล จะต้องทำให้ความซึมเศร้าลดลงเทียบเท่ากับขนาดอิทธิพล 0.5

การกำหนดขนาดกลุ่มตัวอย่าง

- เช่น

ผู้หญิงที่แต่งหน้าจะดูสวยขึ้น

ทดสอบทิศทางเดียว

$\alpha = .05$

ผู้หญิงไม่แต่งหน้า

$\mu = 5$

$\sigma = 2$

ผู้หญิงแต่งหน้า

$\mu = 6$

$\sigma = 2$

ต้องสุ่มผู้หญิงที่แต่งหน้า มากี่คน จึงจะมีกำลังในการทดสอบทางสถิติเท่ากับ .80

การกำหนดขนาดกลุ่มตัวอย่าง

- ตัวอย่าง เช่น

ผู้หญิงที่แต่งหน้าจะดูสวยขึ้น

Null Hypothesis

ทดสอบทางเดียว (α = .05)
↓
หาเปอร์เซนต์ไทล์ที่ 95
↓
$z_{H0} = 1.645$
↓
$\mu_{H0} = 5$

Alternative Hypothesis

กำลัง = .80
↓
หาเปอร์เซนต์ไทล์ที่ 20
↓
$z_{H1} = -0.842$
↓
$\mu_{H1} = 6$

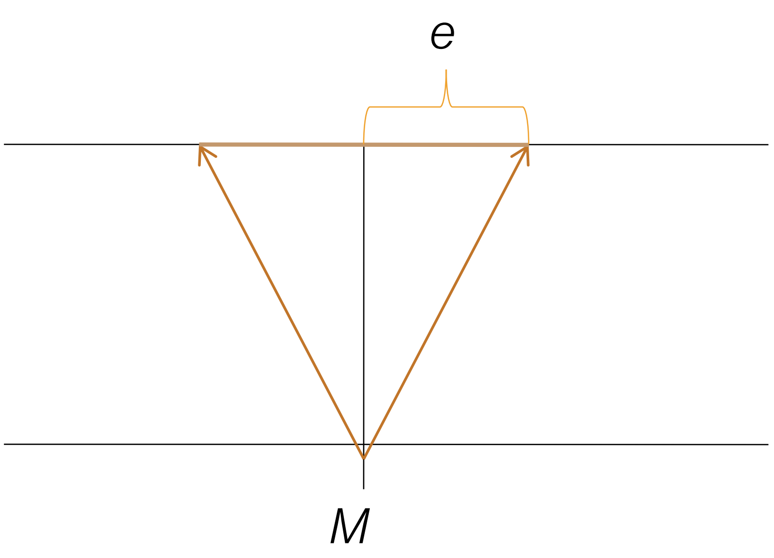
ต้องสุ่มกลุ่มตัวอย่างอย่างน้อย 25 คน จึงจะมีกำลังในการทดสอบทางสถิติ เท่ากับ .80

$$n = \left(\frac{\sigma(z_{H1} - z_{H0})}{\mu_{H0} - \mu_{H1}} \right)^2 = \left(\frac{2(1.645 - (-0.842))}{5 - 6} \right)^2 \approx 25$$

การกำหนดขนาดกลุ่มตัวอย่าง

- การกำหนดขนาดของกลุ่มตัวอย่าง สามารถดูได้อีกรูปแบบหนึ่ง คือ การกำหนดความคาดเคลื่อน (Error; *e*)ในการประมาณค่าพารามิเตอร์
- เช่น ต้องการทำนายรายได้เฉลี่ยของคนไทย ในช่วงเชื่อมั่นระดับ .95 มีความคาดเคลื่อนไม่เกิน 200 บาท
- จากสูตร CI

$$CI_{1-\alpha} = M \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$



การกำหนดขนาดกลุ่มตัวอย่าง

- ความคาดเคลื่อนในที่นี้หมายความว่า

$$e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- แปลงสูตรใหม่ จะได้วิธีการประมาณค่ากลุ่มตัวอย่าง คือ

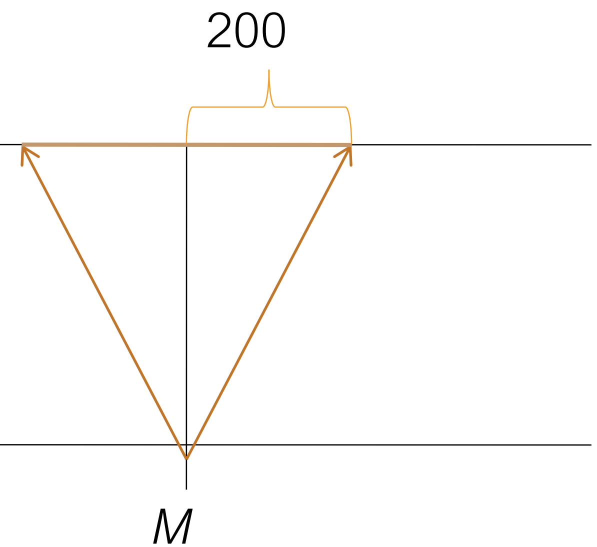
$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$$

การกำหนดขนาดกลุ่มตัวอย่าง

- เช่น ต้องการทำนายรายได้เฉลี่ยของคนไทย ในช่วงเชื่อมั่นระดับ .95 มีความคาดเคลื่อนไม่เกิน 200 บาท ส่วนเบี่ยงเบนมาตรฐานของประชากรเท่ากับ 5,000 บาท

- จากสูตร

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2 = \left(\frac{1.96 \times 5{,}000}{200}\right)^2 = 2{,}401$$



ต้องเก็บข้อมูลจากกลุ่มตัวอย่าง 2,401 คน จึงจะทำให้รายได้เฉลี่ยคาดเคลื่อนไม่เกิน 200 บาท ในช่วงเชื่อมั่นระดับ .95

การกำหนดขนาดกลุ่มตัวอย่าง

- ปัญหาในการประมาณค่า คือ ไม่รู้ส่วนเบี่ยงเบนมาตรฐานของประชากร
- วิธีทางแก้ไข
  1. งานวิจัยเก่า
  2. เก็บข้อมูลก่อนจำนวนหนึ่ง แล้วดูว่าส่วนเบี่ยงเบนมาตรฐานเท่ากับเท่าไร แล้วใช้ค่านั้นแทนค่าส่วนเบี่ยงเบนมาตรฐานของประชากร

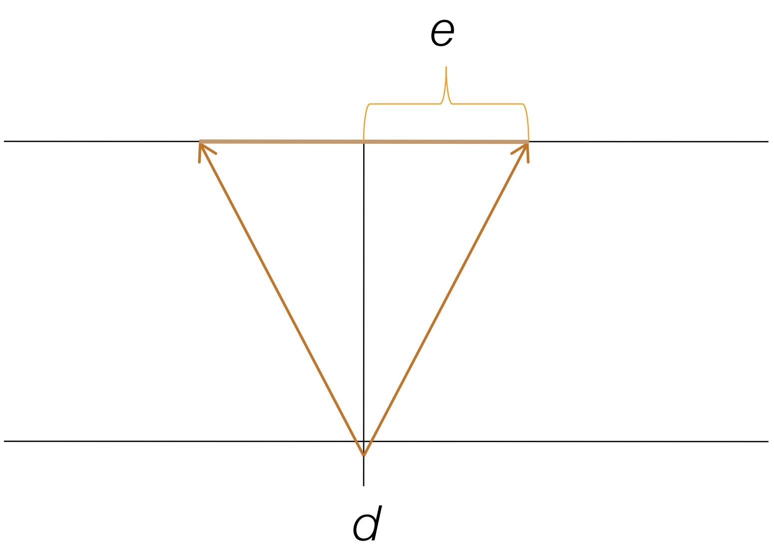
การกำหนดขนาดกลุ่มตัวอย่าง

- การประมาณค่านี้ จะใช้ได้เฉพาะ
  - กลุ่มตัวอย่างแบบไม่จำกัด (Infinite Population) หรือประชากรขนาดใหญ่มากเท่านั้น (ขนาดประมาณ 10 เท่าของกลุ่มตัวอย่างขึ้นไป)
  - สำหรับ Design-based approach การสุ่มต้องเป็นการสุ่มอย่างง่าย (Simple Random Sampling) เท่านั้น ไม่ได้ใช้วิธีการสุ่มรูปแบบอื่น

การกำหนดขนาดกลุ่มตัวอย่าง

- การกำหนดขนาดของกลุ่มตัวอย่าง สามารถดูได้อีกรูปแบบหนึ่ง คือ การกำหนดความคาดเคลื่อน (Error; *e*)ในการประมาณค่าพารามิเตอร์ของขนาดอิทธิพล (δ) ได้ด้วย
- เช่น ต้องการทำนายรายได้เฉลี่ยของคนไทย ในช่วงเชื่อมั่นระดับ .95 มีความคาดเคลื่อนของขนาดอิทธิพลไม่เกิน 0.1
- จากสูตร CI

$$CI_{1-\alpha} = d \pm \frac{z_{\alpha/2}}{\sqrt{n}}$$



การกำหนดขนาดกลุ่มตัวอย่าง

- ความคาดเคลื่อนในที่นี้หมายความว่า

$$e_\delta = \frac{z_{\alpha/2}}{\sqrt{n}}$$

- แปลงสูตรใหม่ จะได้วิธีการประมาณค่ากลุ่มตัวอย่าง คือ

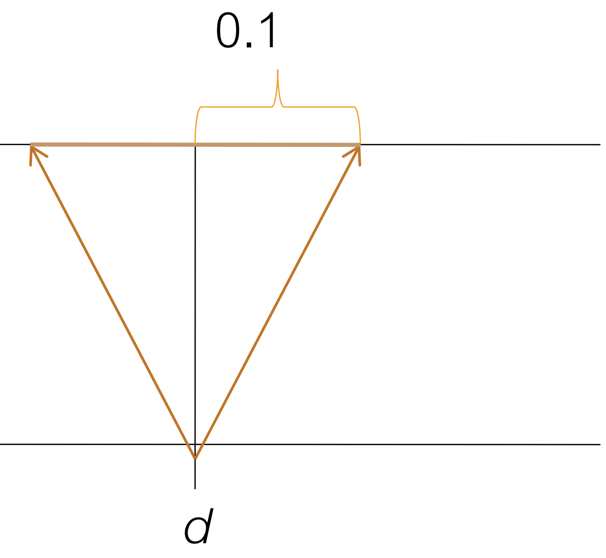
$$n = \left(\frac{z_{\alpha/2}}{e_\delta} \right)^2$$

การกำหนดขนาดกลุ่มตัวอย่าง

- เช่น ต้องการทำนายขนาดอิทธิพลของคอร์สฝึกอบรม ว่ามีขนาดอิทธิพลเท่าไร ในช่วงเชื่อมั่นระดับ .95 มีความคาดเคลื่อนไม่เกิน 0.1
- จากสูตร

$$n = \left(\frac{z_{\alpha/2}}{e_\delta} \right)^2 = \left(\frac{1.96}{0.1} \right)^2 = 384.16$$

ต้องเก็บข้อมูลจากกลุ่มตัวอย่าง 385 คน จึงจะทำให้ขนาดอิทธิพลคาดเคลื่อนไม่เกิน 0.1 ในช่วงเชื่อมั่นระดับ .95



การกำหนดขนาดกลุ่มตัวอย่าง

- การประมาณค่านี้ จะใช้ได้เฉพาะ
  - กลุ่มตัวอย่างแบบไม่จำกัด (Infinite Population) หรือประชากรขนาดใหญ่มากเท่านั้น (ขนาดประมาณ 10 เท่าของกลุ่มตัวอย่างขึ้นไป)
  - สำหรับ Design-based approach การสุ่มต้องเป็นการสุ่มอย่างง่าย (Simple Random Sampling) เท่านั้น ไม่ได้ใช้วิธีการสุ่มรูปแบบอื่น

การกำหนดขนาดกลุ่มตัวอย่าง

- เป้าหมายในการหากลุ่มตัวอย่างทั้ง 3 วิธีแตกต่างกัน

หา *n* เพื่อดูโอกาสในการปฏิเสธสมมติฐานทางสถิติ →	กำหนดกำลังที่ต้องการ
หา *n* เพื่อประมาณค่า μ →	กำหนดความคาดเคลื่อนในการประมาณค่า μ
หา *n* เพื่อประมาณค่า δ →	กำหนดความคาดเคลื่อนในการประมาณค่า δ

การกำหนดขนาดกลุ่มตัวอย่าง

- ถ้าใช้สถิติอ้างอิงอื่น (เช่น ประมาณค่าร้อยละ) วิธีการประมาณค่ากลุ่มตัวอย่างจะเปลี่ยนไป
- ถ้าใช้วิธีการประมาณกลุ่มตัวอย่างหลายสูตร (ใช้หลายวัตถุประสงค์) ให้เลือกจำนวนกลุ่มตัวอย่างสูงสุด

การกำหนดระดับนัยสำคัญ

- การกำหนดระดับนัยสำคัญ ต้องดูความรุนแรงของความผิดพลาดแบบที่ 1 และแบบที่ 2
- สมมติว่า กำหนดให้ความผิดพลาดแบบที่ 1 มี 5 % (α = .05) และความผิดพลาดแบบที่ 2 มี 20 % (กำลัง = .80)
- แสดงว่ายอมให้ความผิดพลาดแบบที่ 2 เกิดขึ้นได้เป็น 4 เท่าของความผิดพลาดแบบที่ 1
- แต่บางครั้งความผิดพลาดแบบที่ 1 รุนแรงมาก เช่น

การกำหนดระดับนัยสำคัญ

- กลุ่มคน 4 คนไปออกค่ายที่ป่าแห่งหนึ่ง มีระดับเกล็ดเลือดเฉลี่ยเท่ากับ 100,000 ชิ้นต่อลูกบาศก์มิลลิเมตร
- แต่คนปกติมีระดับเกล็ดเลือกประมาณ 300,000 ชิ้นต่อลูกบาศก์มิลลิเมตร ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 75,000 ชิ้นต่อลูกบาศก์มิลลิเมตร
- ถ้าคน 4 คนมีเกล็ดเลือดต่ำกว่าปกติ อาจมีแนวโน้มเป็นโรคจากการออกค่าย เช่น ไข้เลือดออก

การกำหนดระดับนัยสำคัญ

- ความผิดพลาดแบบที่ 1 คือ ตัดสินว่าคน 4 คนนี้เป็นโรค ทั้งที่เขาเป็นคนปกติ
- ความผิดพลาดแบบที่ 2 คือ ตัดสินว่าคน 4 คนนี้อาจจะเป็นโรคหรือไม่ก็ได้ ทั้งที่เขาเป็นโรค
- ในที่นี้ ความผิดพลาดแบบที่ 2 อาจมีความเสียหายมาก
- ดังนั้นจึงลดเกณฑ์การควบคุมความผิดพลาดแบบที่ 1 ลง กล่าวคือ ลดระดับนัยสำคัญ เช่น กำหนด $\alpha = .10$ เพื่อลดโอกาสการเกิดความผิดพลาดแบบที่ 2

ข้อผิดพลาดที่มักจะเกิดขึ้นใน Inferential Statistics

- การไม่ถึงระดับนัยสำคัญทางสถิติ แล้วบอกว่า Null Hypothesis เป็นจริง
- การนำค่า p มาใช้ในการบอกขนาดอิทธิพล
- การตีความช่วงเชื่อมั่น ว่าเป็นโอกาสที่ Parameter อยู่ในช่วงดังกล่าว
- การปฏิเสธสมมติฐานทางเดียว แม้ข้อมูลจะอยู่คนละข้าง
- การเกือบถึงระดับนัยสำคัญ (Marginal Significance) เช่น กำหนด $\alpha = .05$ แต่ค่า $p = .06$ นักวิจัยบางคนบอกว่าแตกต่างอย่างมีนัยสำคัญ
  - เป็นประเด็นถกเถียงในวงการวิชาการ
  - สำหรับผม ไม่สมควรทำอย่างยิ่ง

การทดสอบความใกล้เคียง

- การทดสอบว่า Null Hypothesis ไม่จริง เป็นเรื่องที่จริงๆ แล้วดูตลก เพราะว่า Null Hypothesis ไม่เป็นจริงอยู่แล้ว
  - เช่น Null Hypothesis ตั้งว่าประชากรสองกลุ่มมีค่าเฉลี่ยเท่ากัน
  - Null Hypothesis นี้ ไม่เป็นจริงอยู่แล้ว เป็นไปไม่ได้ที่ประชากรสองกลุ่มเท่ากันทุกๆ จุดทศนิยม โดยเฉพาะอย่างยิ่งการวิจัยที่ไม่ใช่การทดลอง (Nonexperimental Research)
- ดังนั้น หากพบความแตกต่างอย่างมีนัยสำคัญ ควรตรวจสอบขนาดอิทธิพลเสมอ

การทดสอบความใกล้เคียง

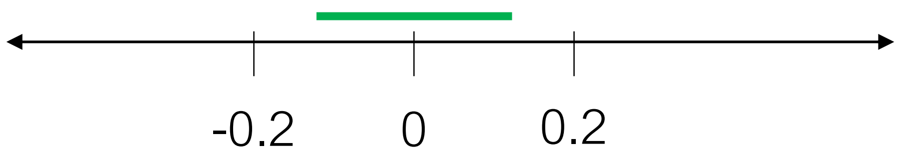
- หากต้องการตรวจสอบว่าค่าเฉลี่ยของสองกลุ่มเท่ากันหรือไม่ การตรวจสอบนี้ไม่ควรทำ เพราะค่าเฉลี่ยของสองกลุ่มไม่มีทางเหมือนกันทุกจุดทศนิยม
- หลีกเลี่ยงไม่ได้ที่จะเจอคำถามวิจัยที่ทดสอบว่าสองกลุ่มไม่แตกต่างกัน
  - การทดลอง ไม่ต้องการให้สองกลุ่มแตกต่างกันก่อนทดลอง
  - การทดสอบข้อตกลงเบื้องต้นก่อนการใช้สถิติ

การทดสอบความใกล้เคียง

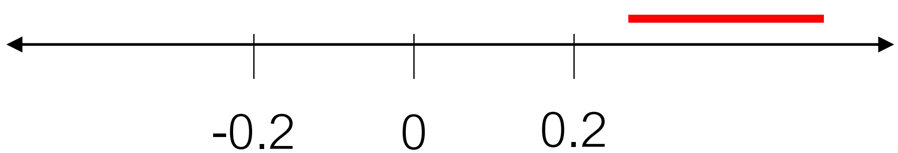
- เปลี่ยนคำถามใหม่ว่าเป็น
  “ใกล้เคียงกันหรือความแตกต่างที่เกิดขึ้นน้อยมากจนราวกับว่าไม่แตกต่างกัน”
- การทดสอบนี้เรียกว่า การทดสอบความใกล้เคียง (Equivalence Testing)
- ผู้วิจัยสามารถกำหนดขอบเขตความใกล้เคียงจาก
  - คะแนนดิบ เช่น IQ แตกต่างไม่เกิน ±5 แต้ม ถือว่าใกล้เคียง
  - ขนาดอิทธิพล เช่น ค่า d แตกต่างไม่เกิน ±0.2 ถือว่าใกล้เคียง

การทดสอบความใกล้เคียง

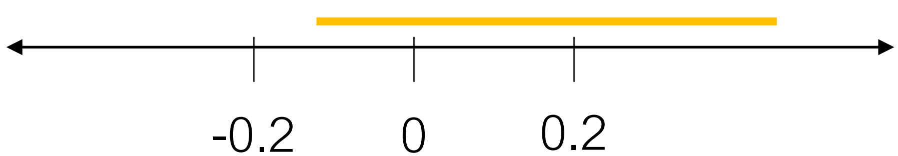
- หลังจากกำหนดขอบเขตใกล้เคียงได้แล้ว ให้ใช้ช่วงเชื่อมั่นในการทดสอบ
- เปรียบเทียบช่วงเชื่อมั่น กับขอบเขตใกล้เคียง
  - ค่าเฉลี่ยใกล้เคียงกัน



  - ค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญในเชิงปฏิบัติ (Practical Significance)



  - ตัดสินใจไม่ได้ ว่าค่าเฉลี่ยใกล้เคียงกันหรือไม่



การทดสอบความใกล้เคียง

- ผลของการตรวจสอบความใกล้เคียง ขึ้นอยู่กับ
  - ตำแหน่งของความแตกต่าง ยิ่งตำแหน่งของค่าเฉลี่ยของกลุ่มตัวอย่าง แตกต่างจากขอบเขตความใกล้เคียง โอกาสเจอผลตัดสินใจไม่ได้ยิ่งน้อย
  - ความกว้างของช่วงเชื่อมั่น ยิ่งช่วงเชื่อมั่นแคบ โอกาสเจอผลตัดสินใจไม่ได้ยิ่งน้อย

การทดสอบความใกล้เคียง

- แต่การตีความจะต้องระมัดระวัง เช่น คนกลุ่มนี้มีความสูงอยู่ในช่วงเชื่อมั่น .95 เท่ากับ (164, 166) หรือขนาดอิทธิพลเท่ากับ
  (-0.1, 0.1)
- สรุปได้ว่า คนกลุ่มนี้มีความสูงเท่ากับคนไทย หรือแทบไม่แตกต่างจากชาวไทย
- แต่ไม่ได้หมายความว่า คนกลุ่มนี้เป็นชาวไทย อาจมีคนกลุ่มอื่นที่มีความสูงเท่ากับคนไทยก็ได้
- หากจะต้องอธิบายว่าคนนี้เป็นชาวไทย จะต้องรวบรวมหลักฐานอื่น จนตัดคำอธิบายที่เป็นไปได้ทั้งหมดออก จนสรุปได้ว่าคนกลุ่มนี้เป็นคนไทย