# การใช้โปรแกรมในการคำนวณ Power และ sample size

วาสิณี จันทร์ธร
สถาบันวิจัยเพื่อการประเมินและออกแบบนโยบาย (RIPED)
มหาวิทยาลัยหอการค้าไทย

October 8, 2022

# Statistical power

- ในการทดสอบสมมติฐานจะมีสองสมมติฐาน คือ สมมติฐานว่าง ($H_0$) และ สมมติฐานทางเลือก ($H_1$) ในการออกแบบที่มีกลุ่มทดลอง 2 กลุ่ม
  - สมมติฐานว่างส่วนใหญ่ระบุว่าค่าเฉลี่ยของกลุ่มประชากรในกลุ่ม treatment และกลุ่ม control ไม่แตกต่างกัน
  - ขณะที่สมมติฐานทางเลือกระบุว่า ค่าเฉลี่ยของกลุ่มประชากรในกลุ่ม treatment และกลุ่ม control แตกต่างกัน
  - ความแตกต่างอาจจะให้ผลการทดลองในเชิงบวก หรือผลการทดลองในเชิงลบ หรืออาจจะเพียงแค่ค่าเฉลี่ยของกลุ่ม treatment ไม่เท่ากับค่าเฉลี่ยของกลุ่ม control
- หลังจากตั้งสมมติฐานไว้อย่างชัดเจนแล้ว และได้รวบรวมและวิเคราะห์ข้อมูลแล้ว ผู้วิจัยต้องตัดสินใจว่ามีหลักฐานเพียงพอหรือไม่ที่จะปฏิเสธสมมติฐานว่างนี้

# Statistical power

- **Power** คือ ความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่าง ($H_0$) เมื่อสมมติฐานทางเลือกเป็นจริง ($H_1$ จริง) ในการศึกษาเปรียบเทียบระหว่างสองกลุ่ม Power คือ โอกาสของการปฏิเสธสมมติฐานว่างที่นิยามว่าทั้งสองกลุ่มมีค่าเฉลี่ยของประชากรร่วมกัน ($H_0 : \mu_1 = \mu_2$) กล่าวคือ โอกาสในการตัดสินใจที่ถูกต้องว่าทั้งสองกลุ่มมีค่าเฉลี่ยของประชากรที่แตกต่างกัน
  - เหตุการ์ที่เรารู้อยู่แล้วว่า $H_1$ เป็นจริง ($H_1 : \mu_1 \neq \mu_2$)
  - ปฏิเสธสมมติฐานว่าง $H_0 : \mu_1 = \mu_2$
  - เมื่อสมมติฐานทางเลือกเป็นจริง $H_1 : \mu_1 \neq \mu_2$
- **Power** ถูกเชื่อมโยงกับการอภิปรายผลการทดสอบสมมติฐานและระดับนัยสำคัญ ($\alpha$) ดังนั้น จึงเป็นสิ่งสำคัญที่จะต้องมีคำจำกัดความที่ชัดเจนของแต่ละคำเหล่านี้ก่อนที่จะดำเนินการในขั้นต่อไป
  - ข้อสังเกต : ในการทดลองแบบสุ่มที่ดำเนินการอย่างสมบูรณ์แบบด้วยการวิเคราะห์ข้อมูลอย่างถูกต้อง Power คือ ความน่าจะเป็นของการค้นพบสาเหตุของผลที่เกิดขึ้นจากการทดลอง (treatment) เมื่อผลกระทบดังกล่าวมีอยู่จริง

# Statistical power

- ระดับนัยสำคัญบ่อยครั้งจะเขียนแทนด้วยสัญลักษณ์ $\alpha$ หมายถึง ความน่าจะเป็นที่ปฏิเสธสมมติฐานว่าง ($H_0$) เมื่อสมมติฐานว่างเป็นจริง ในกรณีนี้เราจะเรียกว่าอัตราความผิดพลาดแบบ Type I error จะเกิดขึ้นเมื่อผู้วิจัยพบความแตกต่างที่มีนัยสำคัญระหว่างสองกลุ่ม แต่แท้จริงแล้วทั้งสองกลุ่มไม่ต่างกัน
- ในทางกลับกันสมมติว่า ความเป็นจริงคือสมมติฐานว่างเป็นเท็จ ความผิดพลาดแบบ Type II error จะเกิดขึ้นเมื่อผู้วิจัยไม่ปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างเป็นเท็จ หรือ ความน่าจะเป็นที่จะคงไว้ซึ่งสมมติฐานว่างที่เป็นเท็จ ซึ่งมักเขียนแทนด้วยสัญลักษณ์ $\beta$

# Statistical power

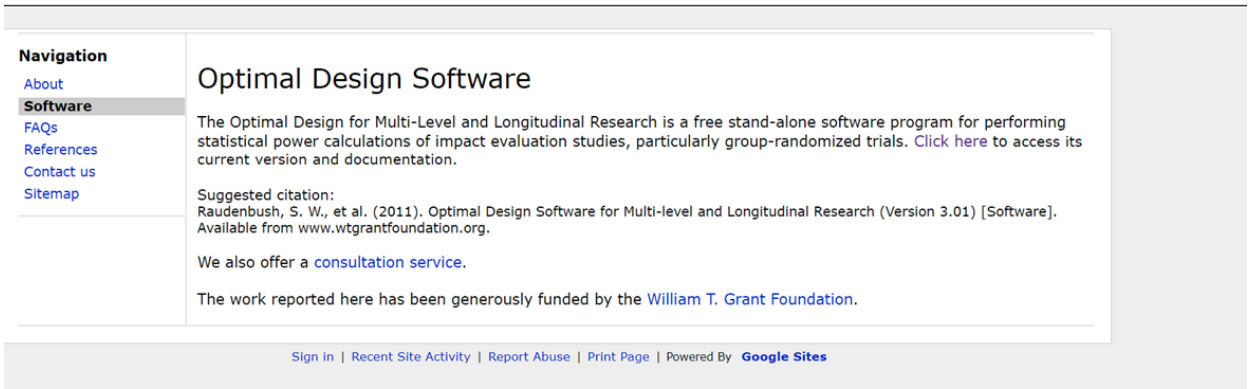
- กรณีที่ผู้วิจัยมองข้ามความแตกต่างที่มีนัยสำคัญของความผิดพลาด 2 ประเภท สามารถแสดงได้จากตาราง Possible errors in hypothesis testing ดังต่อไปนี้

| | ไม่ปฏิเสธสมมติฐานว่าง $H_0$ | ปฏิเสธสมมตฐานว่าง $H_0$ |
|---|---|---|
| สมมติฐานว่าง $H_0$ เป็นจริง | ไม่มีความผิดพลาด (ความน่าจะเป็น = $1-\alpha$) | Type I Error (ความน่าจะเป็น = $\alpha$) |
| สมมติฐานว่าง $H_0$ เป็นเท็จ | Type II Error (ความน่าจะเป็น = $\beta$) | Power ไม่มีความผิดพลาด (ความน่าจะเป็น = $1-\beta$) |

  - นักวิจัยส่วนใหญ่จะยอมรับค่า Power ที่มากกว่าหรือเท่ากับ 0.80 กล่าวคือ เราสามารถระบุความแตกต่างได้อย่างถูกต้อง ระหว่างกลุ่ม treatment และกลุ่ม control ด้วยความน่าจะเป็นที่มากกว่าหรือเท่ากับ 0.80

# Optimal Design Software

- โปรแกรมที่ใช้ในการคำนวณ Power และขนาดตัวอย่าง คือ Optimal Design Sofware หรือ OD (พัฒนาโดย Stephen Raudenbush, Jessaca Spybrook, Howard Bloom, Richard Congdon, Carolyn Hilld, Andres Martineze, October 16, 2011)



# Power and sample size : person randomized trials

- กรณีที่สุ่มระดับบุคคล (person randomized trials)
- แบบจำลอง

$$Y_i = \beta_0 + \beta_1 W_i + r_i; r_i \sim N(0, \sigma^2) \tag{1}$$

  - สำหรับ $i = 1, ..., N$ คนในการศึกษา
  - โดยที่ $Y_i$ เป็นผลลัพธ์ของคนที่ $i$
  - $\beta_0$ เป็นค่าเฉลี่ยของผลลัพธ์
  - $\beta_1$ เป็นผลของการทดลอง (treatment effect)
  - $W_i$ เป็นตัวแปรดัมมี่ที่ระบุว่าเป็นกลุ่ม treatment ถ้ามีค่าเท่ากับ $\frac{1}{2}$ และเป็น control ถ้ามีค่าเท่ากับ $\frac{-1}{2}$
  - $r_i$ เป็นค่าคลาดเคลื่อนแบบสุ่มของคนที่ $i$
  - $\sigma^2$ เป็นค่าความแปรปรวน

- จากการทดสอบสมมติฐาน $H_0 : \hat{\beta}_1 = \bar{Y}_T - \bar{Y}_C = 0$ หรือไม่ โดยใช้สถิติทดสอบ F-Test สำหรับแบบจำลองที่อยู่ในรูปแบบมาตรฐาน เรากำหนดให้ $\sigma^2$=1 จะทำให้ได้ค่าพารามิเตอร์ของ Power ที่สามารถคำนวณขนาดตัวอย่าง ($N$) และขนาดของผลการทดลองที่อยู่ในรูปแบบมาตรฐาน ($\delta$) ได้จากสมการดังต่อไปนี้

$$\lambda = \frac{N\beta_1^2/\sigma^2}{4\sigma^2/\sigma^2} = \frac{N\delta^2}{4} \tag{2}$$

  - เมื่อ $\delta = \frac{\beta_1}{\sqrt{\sigma^2}}$

- ในกรณีที่แบบจำลองมีตัวแปรควบคุม (control variables)

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 X_i + r_i; r_i \sim N(0, \sigma^2) \tag{3}$$

  - โดยที่ $\beta_2$ เป็นค่าสัมประสิทธิ์การถดถอยของตัวแปรควบคุม
  - $X_i$ เป็นตัวแปรควบคุม (control variables)

- สามารถคำนวณขนาดตัวอย่าง ($N$) และขนาดของผลการทดลองที่อยู่ในรูปแบบมาตรฐาน ($\delta$) ได้จากสมการดังต่อไปนี้

$$\lambda = \frac{N\delta^2}{4(1 - \rho_{xy}^2)} \tag{4}$$

  - โดยที่ $\rho_{xy}$ เป็นค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง $X$ และ $Y$

# Example I : Power and sample size

RIPED
Research Institute for Policy Evaluation and Design

- ตัวอย่าง : ทีมนักวิจัยกำลังวางแผนที่จะทำการทดลองเพื่อตรวจสอบว่าการเข้าโรงเรียนทางเลือก (the charter schools) สามารถช่วยให้นักเรียนมีผลสัมฤทธิ์ทางการเรียนดีขึ้นหรือไม่ เมื่อเปรียบเทียบกับการเข้าโรงเรียนรัฐบาล (the local public school)
- สมมติว่ามีนักเรียนสมัครเข้าเรียนโรงเรียนทางเลือกเป็นจำนวนมากเกินกว่าจะรับเข้าได้ เงื่อนไขก็คือนักเรียนทุกคนที่สมัครเข้ามาจะต้องจับฉลากเพื่อให้มีสิทธิ์ได้เข้าเรียน โดยครึ่งหนึ่งจะได้เข้าเรียนในโรงเรียนทางเลือก และอีกครึ่งหนึ่งจะได้เข้าเรียนในโรงเรียนรัฐบาล
- นักวิจัยตั้งสมมติฐานว่า นักเรียนที่ลงทะเบียนเรียนในโรงเรียนทางเลือกจะมีผลสัมฤทธิ์ทางการเรียนมากกว่านักเรียนลงทะเบียนเรียนในโรงเรียนรัฐบาล โดยพวกเขาวางแผนที่จะวัดความสำเร็จโดยใช้แบบทดสอบของ the Iowa Test of Basic Skills (ITBS)

# Example I : Power and sample size

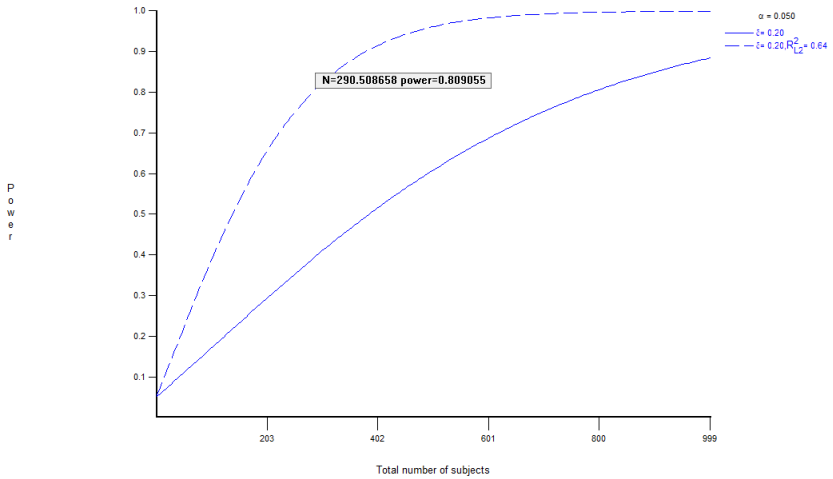
- จากผลการศึกษานำร่อง (pilot tudy) นักวิจัยคาดหวังว่านักเรียนในกลุ่ม treatment จะมีคะแนน 0.25 SD มากกว่านักเรียนในกลุ่ม control สำหรับการทดสอบโดยใช้ ITBS
- นักวิจัยต้องการความเชื่อมั่นทางสถิติ Statistical Power = 0.80
- สมมุติว่าผู้วิจัยตัดสินใจทำการทดสอบ pre-test กับนักเรียนทุกคนก่อนเข้าศึกษา โดยอ้างอิงจากงานวิจัยที่ผ่านมา พวกเขาคาดว่าผลการทดสอบ pre-test จะสามารถอธิบายผลการทดสอบ post-test ได้ถึง 64%
- หลังจากนำข้อมูลการทดสอบ pre-test รวมเข้ามาในการออกแบบแล้ว ควรจะมีจำนวนนักเรียนเท่าไหรในการศึกษาครั้งนี้

# Example I : Power and sample size calculation

- ขั้นตอนการคำนวณ
  - ขั้นที่ 1: เลือก Person randomized trials
  - เลือก single level trials
  - เลือก Power on y-axis power vs. total number of people (N)
  - ขั้นที่ 2 : คลิกที่ $\alpha$ กำหนดให้ alpha = 0.05
  - คลิกที่ $\delta$ กำหนดให้ delta(1) = 0.25
  - คลิกที่ $R^2$ กำหนดให้ r2(2) = 0.64
  - คลิกที่ $\leq X \leq$ กำหนดให้ maximum = 1000
  - ขั้นที่ 3 : คลิกที่เส้นโค้งของกราฟเพื่อหาค่า $N$ ที่ค่า Power = 0.80

# Example I : Power and sample size

- จากกราฟจะได้ว่าในการศึกษาครั้งนี้ ควรจะมีจำนวนนักเรียน 290 คน แบ่งเป็นนักเรียนในกลุ่ม treatment จำนวน 145 คน และนักเรียนในกลุ่ม control จำนวน 145 คน
  - โปรแกรม OD จะให้จำนวน N ทั้งหมดที่รวมทั้ง 2 กลุ่มเข้าไว้ด้วยกัน หมายความว่าหากเราได้ค่า N=290 เราจำเป็นจะต้องนำ N มาหาร 2 ก่อน เพื่อดูว่า N ต่อหนึ่งกลุ่มมีจำนวนเท่าไหร่ ซึ่งในกรณีนี้จะมีจำนวนกลุ่มละ 145 คน
  - เช่นเดียวกัน ในกรณีที่เรามีกลุ่มเปรียบเทียบ 3 กลุ่ม ในแต่ละกลุ่มจะต้องมีจำนวน 145 คน เพราะฉะนั้นจะต้องมีจำนวนตัวอย่างทั้งหมด 3 x 145 = 435 คน



- กรณีที่สุ่มระดับกลุ่ม (cluster randomized trials)
- แบบจำลองระดับบุคคล (The level-1)

$$Y_{ij} = \beta_{0j} + e_{ij}; e_{ij} \sim N(0, \sigma^2) \tag{5}$$

  - สำหรับ $i \in \{1, ..., n\}$ จำนวนคนต่อกลุ่ม และ $j \in \{1, ..., J\}$ จำนวนกลุ่ม
  - โดยที่ $Y_{ij}$ เป็นผลลัพธ์ของคนที่ $i$ ในกลุ่มที่ $j$
  - $\beta_{0j}$ เป็นค่าเฉลี่ยของผลลัพธ์ในกลุ่มที่ $j$
  - $e_{ij}$ เป็นค่าคลาดเคลื่อนแบบสุ่มของคนที่ $i$ ในกลุ่มที่ $j$
  - $\sigma^2$ เป็นค่าความแปรปรวนภายในกลุ่ม

- แบบจำลองระดับกลุ่ม (The level-2)

$$\beta_{oj} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j}; \mu_{0j} \sim N(0, \tau) \tag{6}$$

  - โดยที่ $\gamma_{00}$ เป็นค่าเฉลี่ยทั้งหมดของผลลัพธ์
  - $\gamma_{01}$ เป็นความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม treatment และกลุ่ม control
  - $W_j$ เป็นตัวแปรดัมมี่ที่ระบุว่าเป็นกลุ่ม treatment ถ้ามีค่าเท่ากับ $\frac{1}{2}$ และเป็น control ถ้ามีค่าเท่ากับ $\frac{-1}{2}$
  - $\mu_{0j}$ เป็นค่าคลาดเคลื่อนอย่างสุ่มที่สัมพันธ์กันในแต่ละกลุ่ม
  - $\tau$ เป็นค่าความแปรปรวนระหว่างกลุ่ม
- แทนสมการที่ (6) ลงในสมการที่ (5) จะได้แบบจำลองดังนี้

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j} + e_{ij} \tag{7}$$

  - โดยที่ $\mu_{0j} \sim N(0, \tau)$ และ $e_{ij} \sim N(0, \sigma^2)$

- จากการทดสอบสมมติฐาน $H_0 : \hat{\gamma}_{01} = 0$ หรือไม่ โดยใช้สถิติทดสอบ F-Test สำหรับแบบจำลองที่อยู่ในรูปแบบมาตรฐาน เรากำหนดให้ $\sigma^2$=1 จะทำให้ได้ค่าพารามิเตอร์ของ Power ที่สามารถคำนวณจำนวนตัวอย่างต่อกลุ่ม ($n$) จำนวนกลุ่ม ($J$) และขนาดของผลการทดลองที่อยู่ในรูปแบบมาตรฐาน ($\delta$) ได้จากสมการดังต่อไปนี้

$$\lambda = \frac{J\delta^2}{4(\rho + (1-\rho)/n)} \tag{8}$$

  - เมื่อ $\delta = \frac{\gamma_{01}}{\sqrt{\tau+\sigma^2}}$
  - ค่า intra-class cluster correlation = $\rho = \frac{\tau}{\tau+\sigma^2}$
  - ค่า intra-class cluster correlation คือ ค่าความแปรปรวนที่เป็นผลมาจากความแตกต่างระหว่างกลุ่ม ที่บอกถึงความคล้ายคลึงกันของกลุ่มตัวอย่างในแต่ละกลุ่ม ค่ายิ่งสูงหมายถึงภายในกลุ่มมีความคล้ายคลึงกันมาก

- ในกรณีที่แบบจำลองมีตัวแปรควบคุม (control variables)

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{02} X_j + \mu_{0j} + e_{ij} \tag{9}$$

  - โดยที่ $\gamma_{02}$ เป็นค่าสัมประสิทธิ์การถดถอยในระดับกลุ่มของตัวแปรควบคุม
  - $X_i$ เป็นตัวแปรควบคุม (control variables)
  - $\mu_{0j} \sim N(0, \tau_{|x})$ และ $e_{ij} \sim N(0, \sigma^2)$
  - $\tau_{|x}$ เป็นค่าความแปรปรวนของค่าคลาดเคลื่อนระหว่างกลุ่ม

- สามารถคำนวณจำนวนตัวอย่างต่อกลุ่ม ($n$) จำนวนกลุ่ม ($J$) และขนาดของผลการทดลองที่อยู่ในรูปแบบมาตรฐาน ($\delta$) ได้จากสมการดังต่อไปนี้

$$\lambda = \frac{J\delta^2}{4[(1 - R^2_{Level2})\rho + (1 - \rho)/n]} \tag{10}$$

- โดยที่ $R^2_{Level2}$ เป็นค่า $R^2$ ในระดับกลุ่ม

# Example II : Power and sample size

- ตัวอย่าง : สมมติว่าทีมวิจัยต้องการพัฒนาโปรแกรมใหม่เพื่อช่วยเพิ่มทักษะด้านการรู้หนังสือ (literacy program) สำหรับนักเรียนชั้น ป.3 ผู้ก่อตั้งโปรแกรมใหม่นี้มีวัตถุประสงค์เพื่อให้นักเรียนที่เข้าร่วมโปรแกรมมีผลสัมฤทธิ์ทางการอ่านเพิ่มขึ้น
- นักวิจัยวางแผนที่จะทดสอบนักเรียนที่เข้าร่วมโปรแกรม (กลุ่ม treatment) กับนักเรียนที่ไม่ได้เข้าร่วมโปรแกรม (กลุ่ม control) โดยจะใช้แบบทดสอบมาตรฐานเพื่อตรวจสอบว่านักเรียนที่เข้าร่วมโปรแกรมใหม่นี้ได้คะแนนสูงกว่าหรือไม่
- นักวิจัยสามารถเข้าถึงคะแนนการทดสอบด้านการอ่านช่วงปลายภาคของนักเรียนชั้น ป.3 ในแต่ละโรงเรียนได้ ข้อมูลในอดีตแสดงให้เห็นว่าคะแนนในปีที่ผ่านมาสามารถอธิบายคะแนนการทดสอบในระดับชั้น ป.3 ได้ถึง 49%
- นักวิจัยต้องการออกแบบการสุ่มในระดับโรงเรียน

# Example II : Power and sample size

- จากการศึกษาในอดีต นักวิจัยคาดว่ามีประมาณ 20% ของความแปรปรวนที่ได้เป็นผลมาจากความแตกต่างระหว่างโรงเรียน และสนใจขนาดของผลของการทดลองอย่างน้อยที่ 0.25 SD เพื่อให้เพียงพอต่อค่า Statistical Power = 0.80
- สมมติว่ามีนักเรียน 20 คนที่เต็มใจเข้าร่วมโปรแกรมจากแต่ละโรงเรียน นอกจากนี้นักวิจัยต้องการนำข้อมูลของคะแนนในปีที่ผ่านมาที่สามารถอธิบายคะแนนการทดสอบในระดับชั้น ป.3 ได้ถึง 49% รวมเข้ามาในการออกแบบด้วย
- นักวิจัยจะต้องใช้จำนวนโรงเรียนเท่าไหร่ในการศึกษาครั้งนี้

# Example II : Power and sample size calculation

- ขั้นตอนการคำนวณ
  - ขั้นที่ 1: เลือก Cluster randomized trials with person level outcomes
  - เลือก Cluster randomized trials
  - เลือก Treatment at level 2
  - เลือก Power on y-axis (continuous outcome)
  - เลือก Power vs. number of clusters (J)
  - ขั้นที่ 2 : คลิกที่ $\alpha$ กำหนดให้ alpha = 0.05
  - คลิกที่ $n$ กำหนดให้ n(1) = 20
  - คลิกที่ $\delta$ กำหนดให้ delta(1) = 0.25
  - คลิกที่ $\rho$ กำหนดให้ rho(1) = 0.20
  - คลิกที่ $R^2_{Level2}$ กำหนดให้ $R^2_{Level2}(2)$ = 0.49
  - คลิกที่ $\leq X \leq$ กำหนดให้ maximum = 150
  - ขั้นที่ 3 : คลิกที่เส้นโค้งของกราฟเพื่อหาค่า $N$ ที่ค่า Power = 0.80

# Example II : Power and sample size

- จากกราฟจะได้ว่าในการศึกษาครั้งนี้ ควรจะมีจำนวนโรงเรียน 74 โรงเรียน แบ่งเป็นโรงเรียนในกลุ่ม treatment จำนวน 37 โรงเรียน และโรงเรียนในกลุ่ม control จำนวน 37 โรงเรียน