# การสร้างโมเดล (Model Building)

โมเดลพหุระดับ (Multilevel Modeling)

สันทัด พรประเสริฐมานิต

# โครงร่างการนำเสนอ

- จำนวนอิทธิพลสุ่มที่สามารถประมาณค่าได้
- การสร้างโมเดล (Model Building)
- ปัญหาจากตัวแปรอิสระที่ ICC = 0

# จำนวนอิทธิพลสุ่มที่สามารถประมาณค่าได้

- ในการวิเคราะห์ถดถอย หากต้องการประมาณค่าพารามิเตอร์ $m$ ตัว (เช่น $b_0, b_1, \ldots, b_{m-1}$) จำนวนกลุ่มตัวอย่างที่เก็บได้ ต้องมีขั้นต่ำ $m + 1$ ตัว เพื่อให้ค่าพารามิเตอร์เหล่านั้น สามารถหาค่าได้
- ในการวิเคราะห์พหุระดับ ค่อนข้างซับซ้อน เนื่องจากพารามิเตอร์จะแบ่งเป็น 2 ชุด คือ อิทธิพลคงที่ (Fixed Effects) และอิทธิพลสุ่ม (Random Effects)
- อิทธิพลสุ่ม จะเป็นอิทธิพลที่แต่ละกลุ่มในระดับที่ 2 จะมีค่าเป็นของตนเอง ดังนั้น จำนวนอิทธิพลสุ่มที่สามารถประมาณค่าได้ ขึ้นอยู่กับจำนวนสมาชิกกลุ่ม ถ้าต้องการประมาณค่าอิทธิพลสุ่ม $m$ ตัว (เช่น $u_0, u_1, \ldots, u_{m-1}$) ควรมีกลุ่มที่มีสมาชิกอย่างน้อย $m + 1$ ตัว จำนวนหนึ่ง

# จำนวนอิทธิพลสุ่มที่สามารถประมาณค่าได้

- กล่าวคือ จำนวนสมาชิกกลุ่ม เป็นข้อจำกัดของจำนวนอิทธิพลสุ่มที่สามารถประมาณค่าได้ เช่น
  - การวิเคราะห์คู่ (Dyads) จะมีอิทธิพลสุ่มได้เพียงจุดตัดเท่านั้น $(u_0)$
  - การวิเคราะห์ข้อมูลระยะยาว ที่เก็บเพียงแค่ 3 ช่วงเวลา จะมีอิทธิพลสุ่มได้แค่ จุดตัด $(u_0)$ และความชันแค่เพียง 1 ตัว $(u_1)$
- ส่วนอิทธิพลคงที่ ซึ่งรวมถึงอิทธิพลหลักของตัวแปรอิสระและปฏิสัมพันธ์ระหว่างชั้น จะใช้ข้อมูลจากทุกกลุ่มมาวิเคราะห์พร้อมกัน ทำให้มักไม่มีข้อจำกัดเรื่องจำนวนกลุ่มตัวอย่าง เพราะการวิเคราะห์พหุระดับต้องเก็บข้อมูลจำนวนเยอะระดับหนึ่งอยู่แล้ว

# การสร้างโมเดล

- ในการอธิบายตัวแปรพหุระดับด้วยตัวแปรอิสระต่างๆ ที่อาจมีจำนวนมาก นักวิจัยต้องตัดสินใจในสิ่งต่อไปนี้
  - ให้นำตัวแปรอิสระตัวใดบ้าง เข้าในโมเดล
  - ยอมให้ตัวแปรอิสระใดบ้างมีความชันสุ่ม
  - ให้มีค่าเฉลี่ยกลุ่ม (หรือค่าสถิติอื่น) ของตัวแปรอิสระ เพื่อทำเป็นตัวแปรอิสระระดับกลุ่มหรือไม่
  - ให้มีปฏิสัมพันธ์หรือไม่ เป็นปฏิสัมพันธ์ระหว่างตัวแปรอะไรบ้าง
- ยิ่งเพิ่มอิทธิพลสุ่ม (Random effect) หรือยิ่งเพิ่มปฏิสัมพันธ์ระหว่างตัวแปรอิสระ จะยิ่งทำให้ความซับซ้อนของโมเดลเพิ่มขึ้น

## โมเดลเปล่า

$$Y_{ij} = \beta_{0j} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

ซับซ้อนขึ้น

## RANCOVA

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{1j}X_{2ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$

ซับซ้อนขึ้น

## Random Slope Model

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{1j}X_{2ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$

ซับซ้อนขึ้น

## Cross-level Interaction

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{1j}X_{2ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_{1j} + \gamma_{12}W_{2j} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}W_{1j} + \gamma_{22}W_{2j} + u_{2j}$$

# การสร้างโมเดล

- นึกถึงโมเดลความชันแบบสุ่ม (Random Slope Model) ถ้าตัวแปรอิสระระดับที่ 1 มีอิทธิพลแบบสุ่มทั้งหมด จำนวนพารามิเตอร์เป็นดังนี้

| จำนวนตัวแปร | จำนวนพารามิเตอร์ | รายละเอียด |
|---|---|---|
| 0 | 3 | $\gamma_{00}, \sigma^2, \tau_{00}$ |
| 1 | 6 | $\gamma_{00}, \gamma_{10}, \sigma^2, \tau_{00}, \tau_{10}, \tau_{11}$ |
| 2 | 10 | $\gamma_{00}, \gamma_{10}, \gamma_{20}, \sigma^2, \tau_{00}, \tau_{10}, \tau_{11}, \tau_{20}, \tau_{21}\tau_{22}$ |
| 3 | 15 | $\gamma_{00}, \gamma_{10}, \gamma_{20}, \gamma_{30}, \sigma^2, \tau_{00}, \tau_{10}, \tau_{11}, \tau_{20}, \tau_{21}\tau_{22}, \tau_{30}, \tau_{31}, \tau_{32}, \tau_{33}$ |
| 4 | 21 | $\gamma_{00}, \gamma_{10}, \gamma_{20}, \gamma_{30}, \gamma_{40}, \sigma^2,$ <br> $\tau_{00}, \tau_{10}, \tau_{11}, \tau_{20}, \tau_{21}\tau_{22}, \tau_{30}, \tau_{31}, \tau_{32}, \tau_{33}, \tau_{40}, \tau_{41}, \tau_{42}, \tau_{43}, \tau_{44}$ |
| $q$ | $\frac{q^2 + 5q + 6}{2}$ | จำนวนพารามิเตอร์ เพิ่มขึ้นเป็นเส้นโค้ง ยิ่งทำให้โมเดลซับซ้อนขึ้น |

# การสร้างโมเดล

- ยิ่งประมาณค่าพารามิเตอร์ที่มากขึ้น (เช่น การเพิ่มอิทธิพลสุ่ม การเพิ่มปฏิสัมพันธ์ การเพิ่มตัวแปร) ยิ่งเพิ่มความซับซ้อนของโมเดล
- ความซับซ้อนของโมเดลยิ่งมาก ยิ่งทำให้ผู้วิจัยอธิบายโมเดลได้ยาก และการประมาณค่าผ่านโปรแกรมอาจไม่ลู่เข้าสู่ผลลัพธ์ (Nonconvergence)
    - การวิเคราะห์ความเป็นไปได้สูงสุด (Maximum Likelihood) จะค้นหาค่าพารามิเตอร์ในคอมพิวเตอร์ โดยจะค่อยๆ ปรับค่าของพารามิเตอร์ไปเรื่อยๆ จนกว่าจะเจอค่าที่เป็นไปได้สูงสุด
    - ถ้าค้นไปตั้งนานแล้ว ยังไม่เจอค่าเป็นไปได้สูงสุด จะเรียกว่าโปรแกรมไม่ลู่เข้าสู่ผลลัพธ์ (Nonconvergence)

# การสร้างโมเดล

- ดังนั้น การเพิ่มพารามิเตอร์ จะต้องเป็นการเพิ่มที่สำคัญจริงๆ
- ลักษณะของโมเดลที่ดี จะต้อง
  - อธิบายข้อมูลได้ดี (Fit) ซึ่งหมายความว่า ไม่ได้มีพารามิเตอร์น้อยจนเกินไป จนอธิบายความสัมพันธ์ระหว่างตัวแปรไม่ครบถ้วน
  - ประหยัด (Parsimonious) ไม่ใช้พารามิเตอร์เยอะเกินจำเป็น
- การสร้างโมเดล มีสองวิธี คือ
  - สร้างโมเดลจากทฤษฎีโดยตรง หรือที่เรียกว่าวิธีนิรนัย (deductive)
  - สร้างโมเดลจากข้อมูล หรือที่เรียกว่าวิธีอุปนัย (inductive)

# การสร้างโมเดล

- โมเดลที่สร้างจากทฤษฎีโดยตรง เป็นการจัดวางตัวแปรอิสระ อิทธิพลสุ่ม หรือปฏิสัมพันธ์ระหว่างตัวแปรลงไปในโมเดลตั้งแต่ครั้งแรกเลย เช่น
  - ทดสอบอิทธิพลการดึงดูดจากความเหมือน (Similarity-Attraction Effect) โดยให้ผู้ร่วมการทดลองประเมินภาพที่มีความเข้มของสีผิวแตกต่างกัน (ภาพ:ผู้ร่วมการทดลอง) และเก็บข้อมูลความเข้มของสีผิวของผู้ร่วมการทดลอง
  - สร้างโมเดลที่มีปฏิสัมพันธ์ระหว่างระดับได้เลย

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_{1j} + u_{1j}$$

$Y$ = ความชอบพอต่อคนในรูปภาพ

$X_1$ = ความเข้มของสีผิวของรูปคนในภาพ

$W_1$ = ความเข้มของสีผิวของผู้ร่วมการทดลอง

คาดหวังให้ $\gamma_{11}$ ถึงระดับนัยสำคัญ ไปในทิศทางบวก

- ครูที่สอนแบบเด็กเป็นศูนย์กลาง จะเหมาะสมกับเด็กที่มีความต้องการทางปัญญาสูง แต่การสอนแบบบรรยาย จะเหมาะสมกับเด็กที่มีความต้องการทางปัญญาต่ำ โดยสะท้อนจากผลการสอบเข้าอุดมศึกษาวิชาสังคมศาสตร์ ควบคุมเพศและคะแนนวิชาสังคมศาสตร์เดิม

$$Y_{ij} = \beta_{0j} + \beta_{1j}\left(X_{1ij} - \bar{X}_{1.j}\right) + \beta_{2j}\left(X_{2ij} - \bar{X}_{2.j}\right) + \beta_{3j}X_{3ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}\bar{X}_{1.j} + \gamma_{02}\bar{X}_{2.j} + \gamma_{03}W_{1j} + \gamma_{04}W_{1j}\bar{X}_{1.j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{13}W_{1j} + u_{1j}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$

$Y$ = คะแนนสังคมศาสตร์ใหม่

$X_1$ = คะแนนความต้องการทางปัญญา

$X_2$ = คะแนนวิชาสังคมศาสตร์เดิม

$X_3$ = เพศของนักเรียน

$W_1$ = วิธีการสอนของครูในชั้นเรียน (1 = เด็กเป็นศูนย์กลาง, 0 = บรรยาย)

ในการทดสอบทฤษฎี จะคาดหวังให้

- $\gamma_{13}$ ไปทางบวก ซึ่งแสดงว่าเด็กที่มีความต้องการทางปัญญาสูงภายในห้องเรียน การเรียนแบบเด็กเป็นศูนย์กลาง จะยิ่งได้คะแนนสูงกว่าการเรียนแบบบรรยาย
- $\gamma_{04}$ ไปทางบวก ซึ่งแสดงว่าห้องที่ความต้องการทางปัญญาของเด็กเฉลี่ยสูง จะทำให้การเรียนแบบเด็กเป็นศูนย์กลาง จะยิ่งได้คะแนนสูงกว่าการเรียนแบบบรรยาย

# การสร้างโมเดล

- ในกรณีที่ผู้วิจัยทฤษฎีมีหลายทฤษฎี แล้วต้องการเปรียบเทียบทฤษฎีดังกล่าว นักวิจัยจะสร้างสองโมเดลจากสองทฤษฎี เพื่อนำมาเปรียบเทียบกัน โดย
  - ถ้าเป็นโมเดลซ้อนกัน (Nested Model) ใช้การทดสอบเปรียบสัดส่วนความเป็นไปได้ (Likelihood ratio test)
  - หากเปรียบเทียบโมเดลที่ไม่ซ้อนกัน (Nonnested Model) ให้ใช้การทดสอบของ Vuong (1989) ซึ่งสามารถอ่านรายละเอียดได้ที่ Merkle, You, & Preacher (2016) ใน R สามารถใช้ package `nonnest2` ได้

# การสร้างโมเดล

- การใช้วิธีนิรนัย ทฤษฎีของคุณจะต้องแน่นหนาพอสมควร ว่าตัวแปรอะไรมีผล มีผลอย่างไร ไปในทิศทางใด ดังเช่นตัวอย่างข้างต้น
- แต่หากทฤษฎีของคุณอ่อน เช่น ลักษณะของคนมีผลต่อตัวแปรตาม หรือลักษณะของสถานการณ์มีผลต่อตัวแปรตาม ซึ่งตัวแปรอิสระที่จะมีผลต่อตัวแปรตามมีจำนวนมาก และไม่รู้ทิศทางที่ชัดเจน ควรจะใช้วิธีอุปนัย

# การสร้างโมเดล

- การสร้างโมเดลด้วยวิธีอุปนัย (Subjective) เหมือนเป็นการสร้างทฤษฎีจากข้อมูลที่มี มีวิธีการ 2 วิธี คือ
  - วิธีการค่อยๆ สร้างโมเดล (Build-up Strategy)
  - วิธีการตัดโมเดล (Tear-down Strategy)

# วิธีการค่อยๆ สร้างโมเดล

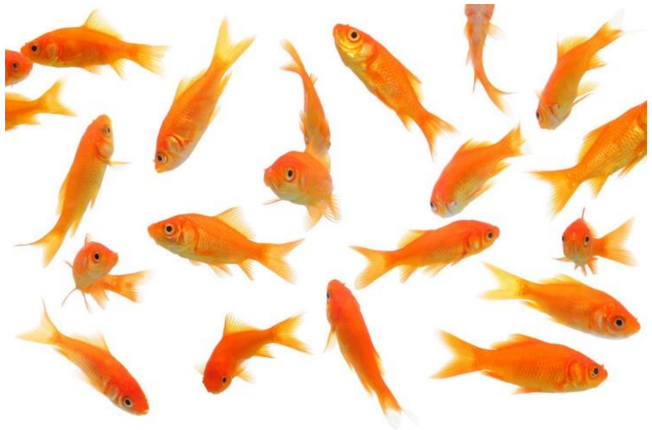
| | |
|---|---|
| โมเดลเปล่า (Null Model) | คำนวณ ICC ว่าตัวแปรนี้เหมาะสมที่จะใช้ MLM หรือไม่ |
| ใส่ตัวแปรอิสระ L1 | ตัวแปรไหนไม่ sig ให้นำออกก่อน |
| ใส่ปฏิสัมพันธ์ระหว่างตัวแปรอิสระ L1 | แม้อิทธิพลหลักของตัวแปรอิสระจะไม่ sig แต่ปฏิสัมพันธ์อาจ sig ได้ ปฏิสัมพันธ์ใดที่ไม่ sig ให้นำออกก่อน |
| ใส่ตัวแปรอิสระ L2 | ตัวแปรไหนไม่ sig ให้นำออกก่อน |
| ใส่ปฏิสัมพันธ์ระหว่างตัวแปรอิสระ L2 | แม้อิทธิพลหลักของตัวแปรอิสระจะไม่ sig แต่ปฏิสัมพันธ์อาจ sig ได้ ปฏิสัมพันธ์ใดที่ไม่ sig ให้นำออกก่อน |
| ทดสอบอิทธิพลสุ่มของตัวแปรอิสระ L1 ทีละตัวแปร | สามารถทดสอบได้ทั้งอิทธิพลหลักและปฏิสัมพันธ์ของตัวแปรอิสระ L1 แม้อิทธิพลจะไม่ sig แต่อาจมีความชันที่แตกต่างระหว่างกลุ่มได้ |
| ทำนายความชันสุ่มของตัวแปรอิสระ L1 ด้วยตัวแปรอิสระ L2 (ปฏิสัมพันธ์ต่างระดับ) | จะเกิดขึ้นกับอิทธิพลของตัวแปรอิสระ L1 ที่มีความชันแบบสุ่มเท่านั้น |

# การสร้างโมเดล

- วิธีการตัดโมเดล (Tear-down Strategy) เป็นวิธีที่ตรงกันข้ามกับวิธีค่อยๆ สร้างโมเดล ที่ใส่ตัวแปรทุกตัว ใส่ปฏิสัมพันธ์ที่เป็นไปได้ทุกตัว แล้วค่อยๆ ลดอิทธิพลที่ไม่ถึงระดับนัยสำคัญลงทีละตัว
- วิธีค่อยๆ สร้างโมเดลและวิธีการตัดโมเดล มักได้โมเดลสุดท้ายที่แตกต่างกัน
- วิธีการตัดโมเดลมักไม่ค่อยได้รับความนิยม เพราะ โมเดลที่ใส่อิทธิพลสุ่มทุกตัวมักจะประมาณค่าไม่สำเร็จ (Nonconvergence)

# การสร้างโมเดล

ทำนายปริมาณอาหารที่ปลาทองกิน



ตัวแปรต้นระดับปลา:

ความยาวปลา,

สีของปลา

ตัวแปรต้นระดับตู้ปลา:

ปริมาตรตู้ปลา,

จำนวนปลาในตู้,

ค่าเฉลี่ยความยาวปลา,

สัดส่วนปลาสีทอง

# ทดสอบตัวแปรอิสระระดับที่ 1

Build-up Strategy

`consume ~ 1 + (1|groupid)`

**โมเดล 0**

IV: ไม่มี

| $f_1$ | $f_2$ | $v$ | $m$ |
|---|---|---|---|
| .00 | .00 | .00 | .49 |

$p < .001$

$p = .065$

**โมเดล 1**

L1 IV: ความยาว (GMC)

`consume ~ 1 + difflengthx + (1|groupid)`

| $f_1$ | $f_2$ | $v$ | $m$ |
|---|---|---|---|
| .32 | .00 | .00 | .55 |

กรณีที่ *m* เพิ่มขึ้น จะอภิปรายในภายหลัง

**โมเดล 2**

L1 IV: สีทอง (GMC)

`consume ~ 1 + diffgold + (1|groupid)`

| $f_1$ | $f_2$ | $v$ | $m$ |
|---|---|---|---|
| .00 | .00 | .00 | .49 |

# ทดสอบปฏิสัมพันธ์ระหว่างตัวแปรอิสระระดับที่ 1

Build-up Strategy

```
consume ~ 1 + difflengthx + diffgold
+ (1|groupid)
```

โมเดลเปรียบเทียบ ต้องใช้โมเดลที่มีอิทธิพลหลักครบ เพื่อให้เปรียบเทียบปฏิสัมพันธ์เพียงอย่างเดียว

**โมเดล 3**

L1 IV: ความยาว (GMC), สีทอง (GMC)

$p$ = .376 ปฏิสัมพันธ์ไม่ sig กลับไปใช้โมเดล 1

**โมเดล 4**

L1 IV: ความยาว (GMC), สีทอง (GMC), ปฏิสัมพันธ์ของทั้งสองตัว

```
consume ~ 1 + difflengthx + diffgold
+ difflengthx:diffgold + (1|groupid)
```

# ทดสอบตัวแปรอิสระระดับที่ 2

**Build-up Strategy**

**โมเดล 1**

L1 IV: ความยาว (GMC)

```
1 + difflengthx
+ (1|groupid)
```

$p < .001$ → **โมเดล 5**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

```
1 + difflengthx
+ volume
+ (1|groupid)
```

$p = .355$ → **โมเดล 6**

L1 IV: ความยาว (GMC)

L2 IV: จำนวนปลา

```
1 + difflengthx
+ nfish
+ (1|groupid)
```

$p < .001$ → **โมเดล 7**

L1 IV: ความยาว (GMC)

L2 IV: ความยาว (M)

```
1 + difflengthx
+ avelengthx
+ (1|groupid)
```

$p = .315$ → **โมเดล 8**

L1 IV: ความยาว (GMC)

L2 IV: สีทอง (M)

```
1 + difflengthx
+ avegold
+ (1|groupid)
```

**โมเดล 9**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

```
1 + difflengthx
+ avelengthx
+ volume
+ (1|groupid)
```

# ทดสอบปฏิสัมพันธ์ระหว่างตัวแปรอิสระระดับที่ 2

**Build-up Strategy**

**โมเดล 9**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

```
1 + difflengthx
+ avelengthx
+ volume
+ (1|groupid)
```

*p* = .397

**โมเดล 10**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

ปริมาตร x ความยาว (M)

```
1 + difflengthx
+ avelengthx
+ volume
+ avelengthx:volume
+ (1|groupid)
```

## ทดสอบปฏิสัมพันธ์ระหว่างตัวแปรอิสระระดับที่ 2

**Build-up Strategy**

โมเดลเปรียบเทียบ ต้องใช้โมเดลที่มีอิทธิพลหลักครบ เพื่อให้เปรียบเทียบปฏิสัมพันธ์เพียงอย่างเดียว

**โมเดล 11**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร,
จำนวนปลา,
ความยาว (M)

```
1 + difflengthx
+ avelengthx
+ volume
+ nfish
+ (1|groupid)
```

*p* = .079

*p* = .852

ปฏิสัมพันธ์ไม่ sig กลับไปใช้โมเดล 9

**โมเดล 12**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร,
จำนวนปลา,
ความยาว (M)
ความยาว (M) x จำนวนปลา

```
1 + difflengthx
+ avelengthx
+ volume
+ nfish
+ avelengthx*nfish
+ (1|groupid)
```

**โมเดล 13**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร,
จำนวนปลา,
ความยาว (M)
ปริมาตร x จำนวนปลา

```
1 + difflengthx
+ avelengthx
+ volume
+ nfish
+ volume*nfish
+ (1|groupid)
```

# ทดสอบปฏิสัมพันธ์ระหว่างตัวแปรอิสระระดับที่ 2

**Build-up Strategy**

โมเดลเปรียบเทียบ ต้องใช้โมเดลที่มีอิทธิพลหลักครบ เพื่อให้เปรียบเทียบปฏิสัมพันธ์เพียงอย่างเดียว

**โมเดล 14**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร,
ความยาว (M),
สีปลา (M)

```
1 + difflengthx
+ volume
+ avelengthx
+ avegold
+ (1|groupid)
```

$p = .144$

$p = .130$

ปฏิสัมพันธ์ไม่ sig กลับไปใช้โมเดล 9

**โมเดล 15**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร,
ความยาว (M),
สีปลา (M),
ความยาว (M) x สีปลา (M)

```
1 + difflengthx
+ volume
+ avelengthx
+ avegold
+ avelengthx*avegold
+ (1|groupid)
```

**โมเดล 16**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร,
ความยาว (M),
สีปลา (M),
ปริมาตร x สีปลา (M)

```
1 + difflengthx
+ volume
+ avelengthx
+ avegold
+ volume*avegold
+ (1|groupid)
```

# ทดสอบปฏิสัมพันธ์ระหว่างตัวแปรอิสระระดับที่ 2

Build-up Strategy

โมเดลเปรียบเทียบ ต้องใช้โมเดลที่มีอิทธิพลหลักครบ เพื่อให้เปรียบเทียบปฏิสัมพันธ์เพียงอย่างเดียว

**โมเดล 17**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร,
จำนวนปลา,
ความยาว (M),
สีปลา (M)

```
1 + difflengthx
+ volume
+ nfish
+ avelengthx
+ avegold
+ (1|groupid)
```

$p$ = .813

ปฏิสัมพันธ์ไม่ sig กลับไปใช้โมเดล 9

**โมเดล 18**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร,
จำนวนปลา,
ความยาว (M),
สีปลา (M)
จำนวนปลา x สีปลา (M)

```
1 + difflengthx
+ volume
+ nfish
+ avelengthx
+ avegold
+ nfish*avegold
+ (1|groupid)
```

# ทดสอบความชันสุ่ม

**Build-up Strategy**

**โมเดล 9**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

```
1 + difflengthx
+ avelengthx
+ volume
+ (1|groupid)
```

$p < .001$

**โมเดล 19**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

ชันสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ avelengthx
+ volume
+ (1 + difflengthx
|groupid)
```

# ทดสอบความชันสุ่ม

Build-up Strategy

ถ้าต้องการเปรียบเทียบความชันสุ่มของตัวแปรใด โมเดลที่เปรียบเทียบต้องใส่อิทธิพลของตัวแปรนั้นด้วย

**โมเดล 20**
L1 IV: ความยาว (GMC)
สีปลา (GMC)
L2 IV: ปริมาตร
ความยาว (M)

```
1 + difflengthx
+ diffgold
+ avelengthx
+ volume
+ (1|groupid)
```

$p$ = .863

ปฏิสัมพันธ์ไม่ sig กลับไปใช้โมเดล 19

**โมเดล 21**
L1 IV: ความยาว (GMC)
สีปลา (GMC)
L2 IV: ปริมาตร
ความยาว (M)
ชันสุ่ม: สีปลา (GMC)

```
1 + difflengthx
+ diffgold
+ avelengthx
+ volume
+ (1 + diffgold
|groupid)
```

# ทดสอบปฏิสัมพันธ์ระหว่างชั้น

**Build-up Strategy**

**โมเดล 19**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

ชันสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ avelengthx
+ volume
+ (1 + difflengthx
|groupid)
```

*p* = .304

*p* = .021

**โมเดล 22**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

L1xL2: ความยาว (GMC) x ความยาว (M)

ชันสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ avelengthx
+ volume
+ difflengthx:avelengthx
+ (1 + difflengthx
|groupid)
```

**โมเดล 23**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

L1xL2: ความยาว (GMC) x ปริมาตร

ชันสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ avelengthx
+ volume
+ difflengthx:volume
+ (1 + difflengthx
|groupid)
```

# ทดสอบปฏิสัมพันธ์ระหว่างชั้น

**Build-up Strategy**

โมเดลเปรียบเทียบ ต้องใช้โมเดลที่มีอิทธิพลหลักครบ เพื่อให้เปรียบเทียบปฏิสัมพันธ์เพียงอย่างเดียว

**โมเดล 24**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

จำนวนปลา

ความยาว (M)

ชั้นสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ avelengthx
+ volume
+ nfish
+ (1 + difflengthx
|groupid)
```

*p* = .081 ปฏิสัมพันธ์ไม่ sig กลับไปใช้โมเดล 23

**โมเดล 25**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

จำนวนปลา

ความยาว (M)

L1xL2: ความยาว (GMC) x จำนวนปลา

ชั้นสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ avelengthx
+ volume
+ nfish
+ difflengthx:nfish
+ (1 + difflengthx
|groupid)
```

# ทดสอบปฏิสัมพันธ์ระหว่างชั้น

Build-up Strategy

โมเดลเปรียบเทียบ ต้องใช้โมเดลที่มีอิทธิพลหลักครบ เพื่อให้เปรียบเทียบปฏิสัมพันธ์เพียงอย่างเดียว

**โมเดล 26**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

สีปลา (M)

ชั้นสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ avelengthx
+ volume
+ avegold
+ (1 + difflengthx
|groupid)
```

*p* = .077   ปฏิสัมพันธ์ไม่ sig กลับไปใช้โมเดล 23

**โมเดล 27**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

สีปลา (M)

L1xL2: ความยาว (GMC) xสีปลา (M)

ชั้นสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ avelengthx
+ volume
+ avegold
+ difflengthx:avegold
+ (1 + difflengthx
|groupid)
```

# ลดปฏิสัมพันธ์ระหว่างชั้น

Tear-down Strategy

**โมเดล 0**

L1 IV: ความยาว (GMC)

สีปลา (GMC)

L2 IV: ความยาว (M)

สีปลา (M)

ปริมาตร,

จำนวนปลา

L1xL2: ความยาว (GMC) x ความยาว (M)

ความยาว (GMC) x สีปลา (M)

ความยาว (GMC) x ปริมาตร

ความยาว (GMC) x จำนวนปลา

สีปลา (GMC) x ความยาว (M)

สีปลา (GMC) x สีปลา (M)

สีปลา (GMC) x ปริมาตร

สีปลา (GMC) x จำนวนปลา

ชันสุ่ม: ความยาว (GMC)

สีปลา (GMC)

```
1 + difflengthx
+ diffgold
+ avelengthx
+ avegold
+ volume
+ nfish
+ difflengthx:avelengthx
+ difflengthx:avegold
+ difflengthx:volume
+ difflengthx:nfish
+ diffgold:avelengthx
+ diffgold:avegold
+ diffgold:volume
+ diffgold:nfish
+ (1 + difflengthx
+ diffgold
|groupid)
```

หมายเหตุ

ผมหลีกเลี่ยงการใส่ปฏิสัมพันธ์ระหว่างตัวแปรอิสระระดับที่ 1 กันเอง และตัวแปรอิสระระดับที่ 2 กันเอง เนื่องจากความซับซ้อนมากเกินไป

# ลดปฏิสัมพันธ์ระหว่างชั้น

**Tear-down Strategy**

**โมเดล 0**

L1 IV: ความยาว (GMC)
สีปลา (GMC)
L2 IV: ความยาว (M)
สีปลา (M)
ปริมาตร,
จำนวนปลา
L1xL2: ความยาว (GMC) x ความยาว (M)
ความยาว (GMC) x สีปลา (M)
ความยาว (GMC) x ปริมาตร
ความยาว (GMC) x จำนวนปลา
สีปลา (GMC) x ความยาว (M)
สีปลา (GMC) x สีปลา (M)
สีปลา (GMC) x ปริมาตร
สีปลา (GMC) x จำนวนปลา
ชันสุ่ม: ความยาว (GMC)
สีปลา (GMC)

$p$ = .910 → **โมเดล 1**: ~~L1xL2: ความยาว (GMC) x ความยาว (M)~~

$p$ = .059 → **โมเดล 2**: ~~L1xL2: ความยาว (GMC) x สีปลา (M)~~

$p$ = .106 → **โมเดล 3**: ~~L1xL2: ความยาว (GMC) x ปริมาตร~~

$p$ = .963 → **โมเดล 4**: ~~L1xL2: ความยาว (GMC) x จำนวนปลา~~

$p$ = .972 → **โมเดล 5**: ~~L1xL2: สีปลา (GMC) x ความยาว (M)~~

$p$ = .990 → **โมเดล 6**: ~~L1xL2: สีปลา (GMC) x สีปลา (M)~~

$p$ = .580 → **โมเดล 7**: ~~L1xL2: สีปลา (GMC) x ปริมาตร~~

$p$ = .278 → **โมเดล 8**: ~~L1xL2: สีปลา (GMC) x จำนวนปลา~~

จากผลการวิเคราะห์นี้ ควรลดปฏิสัมพันธ์ระหว่างชั้นทั้งหมดออกจากโมเดล เป็นโมเดล 9 ในหน้าถัดไป

# ลดความชันสุ่ม

**Tear-down Strategy**

**โมเดล 9**

L1 IV: ความยาว (GMC)

สีปลา (GMC)

L2 IV: ความยาว (M)

สีปลา (M)

ปริมาตร,

จำนวนปลา

ชันสุ่ม: ความยาว (GMC)

สีปลา (GMC)

$p < .001$ →

**โมเดล 10**

~~ชันสุ่ม: ความยาว (GMC)~~

$p = .736$ →

**โมเดล 11**

~~ชันสุ่ม: สีปลา (GMC)~~

ลดความชันสุ่มเฉพาะสีปลา (GMC) ดังนั้นจึงใช้โมเดล 11 ต่อไป

# ลดอิทธิพลตัวแปรอิสระทั้ง L1 และ L2

Tear-down Strategy

**โมเดล 11**
L1 IV: ความยาว (GMC)
สีปลา (GMC)
L2 IV: ความยาว (M)
สีปลา (M)
ปริมาตร,
จำนวนปลา
ชันสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ diffgold
+ avelengthx
+ avegold
+ volume
+ nfish
+ (1 + difflengthx
|groupid)
```

$p = .020$ → **โมเดล 12**
~~L1 IV: สีปลา (GMC)~~

$p < .001$ → **โมเดล 13**
~~L2 IV: ความยาว (M)~~

$p = .009$ → **โมเดล 14**
~~L2 IV: สีปลา (M)~~

$p < .001$ → **โมเดล 15**
~~L2 IV: ปริมาตร~~

$p < .001$ → **โมเดล 16**
~~L2 IV: จำนวนปลา~~

ไม่ต้องลดความยาว (GMC)
เนื่องจากมีความชันแบบสุ่มอยู่

## Build-up Strategy

**โมเดล 23**

L1 IV: ความยาว (GMC)

L2 IV: ปริมาตร

ความยาว (M)

L1xL2: ความยาว (GMC) x ปริมาตร

ชันสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ avelengthx
+ volume
+ difflengthx:volume
+ (1 + difflengthx
|groupid)
```

## Tear-down Strategy

**โมเดล 11**

L1 IV: ความยาว (GMC)

สีปลา (GMC)

L2 IV: ความยาว (M)

สีปลา (M)

ปริมาตร,

จำนวนปลา

ชันสุ่ม: ความยาว (GMC)

```
1 + difflengthx
+ diffgold
+ avelengthx
+ avegold
+ volume
+ nfish
+ (1 + difflengthx
|groupid)
```

โมเดลผลลัพธ์ไม่เหมือนกัน

# การสร้างโมเดล

- จากตัวอย่างที่ผ่านจะแสดงให้เห็นว่า ผลที่ได้จากวิธีค่อยๆ สร้างโมเดลและวิธีค่อยๆ ลดโมเดลออกมาแตกต่างกันอย่างสิ้นเชิง
- หากท่านหลีกเลี่ยงไม่ได้ ต้องค้นหาโมเดลจริงๆ ให้ปฏิบัติดังต่อไปนี้
  - หาจำนวนกลุ่มตัวอย่างให้มากที่สุด โดยเฉพาะอย่างยิ่งในระดับที่ 2 เพื่อให้ได้ค่าพารามิเตอร์ที่เสถียร และการตัดสินใจแต่ละจุดสามารถทำได้ถูกต้อง
  - ในการรายงานผล ให้บอกวิธีการสร้างโมเดลให้ชัดเจน เพราะวิธีการสร้างโมเดลมีผลต่อผลลัพธ์ที่ได้
  - ผลที่ได้ควรมีการสร้างสมมติฐานทางทฤษฎี และมีการวิเคราะห์ซ้ำ (Cross Validation) เพื่อสร้างองค์ความรู้ใหม่

# การสร้างโมเดล

- ในบางครั้ง ผู้วิจัยมีทฤษฎีแล้วบางส่วน แต่ต้องการศึกษาตัวแปรอื่นเพิ่มเติมจากทฤษฎีที่ได้มา จึงสร้างโมเดลจากทฤษฎีก่อน แล้วใช้วิธีอุปนัยในการหาข้อมูลเพิ่มเติม

# การสร้างโมเดล

สมมติฐานงานวิจัย : ทดสอบอคติของผู้สัมภาษณ์ ว่าผู้สัมภาษณ์ประเมินผู้สมัครเพศเดียวกับผู้สัมภาษณ์สูงขึ้นหรือไม่

ตัวแปรต้นระดับผู้สมัคร:

เพศ, เชาวน์ปัญญา, ประสบการณ์ทำงาน (ปี), เงินเดือนที่คาดหวัง (คะแนนมาตรฐานเทียบกับ *M*, *SD* ของตลาดในตำแหน่งเดียวกัน)

ตัวแปรต้นระดับผู้สัมภาษณ์:

เพศ, ประสบการณ์ทำงานสัมภาษณ์ (ปี)

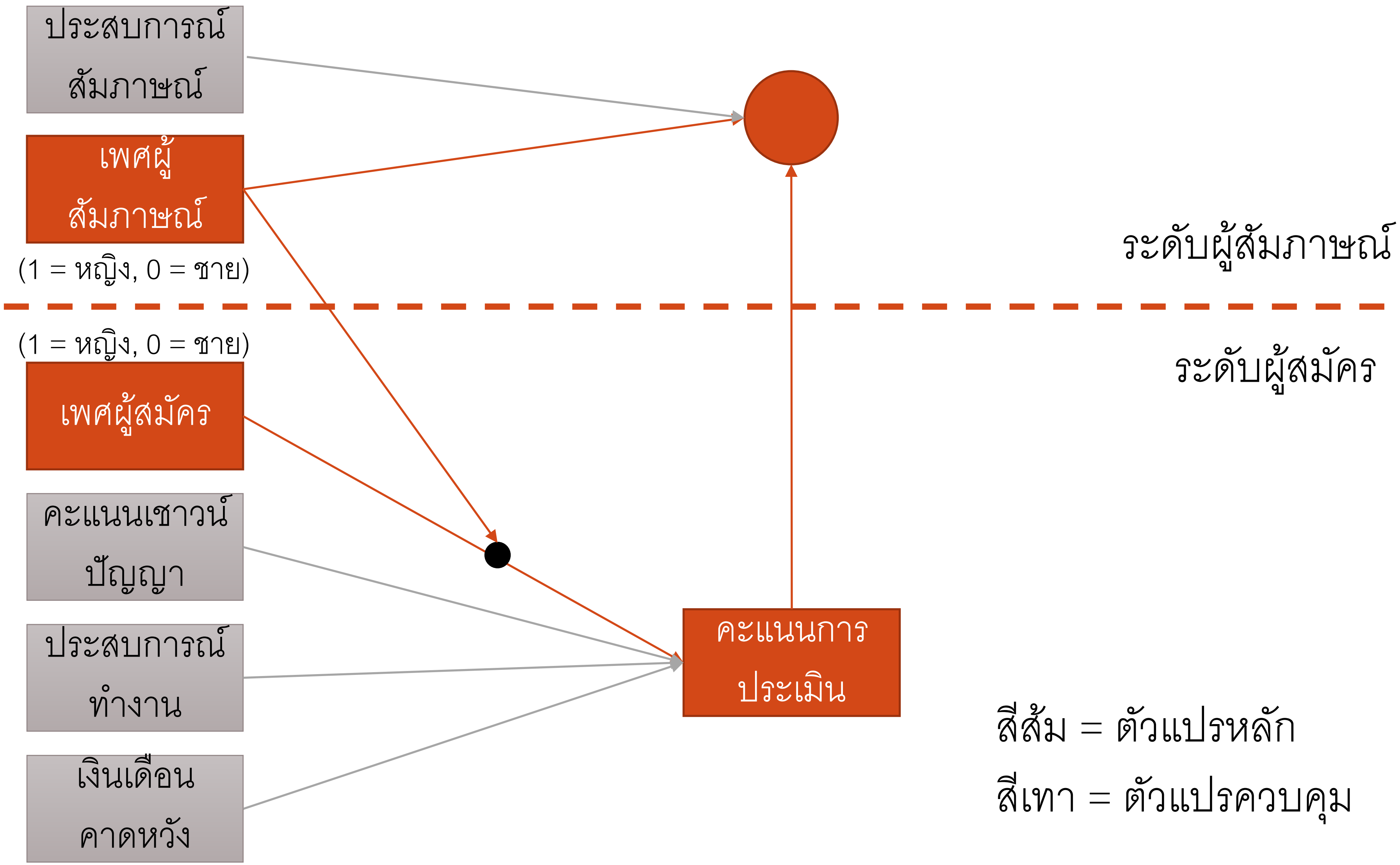
# การสร้างโมเดล

- การทดสอบอิทธิพลนี้ จะเป็นการทดลอง ที่ผู้สมัครจะถูกสุ่มหาผู้สัมภาษณ์ โดยกำหนดโควต้าเพียงผู้สัมภาษณ์แต่ละคนจะสัมภาษณ์ผู้สมัครชาย 5 คน และหญิง 5 คน
- แสดงว่าคุณลักษณะของผู้สมัครที่ผู้สัมภาษณ์แต่ละคนสัมภาษณ์จะไม่แตกต่างกันอย่างเป็นระบบ (แตกต่างกันจากการสุ่มเท่านั้น)
- กล่าวคือ ค่าคาดหวังของการกระจายของประสบการณ์ทำงานของผู้สมัคร, คะแนนเชาวน์ปัญญาของผู้สมัคร, และเงินเดือนคาดหวังของผู้สมัคร จะเท่ากับ 0

# การสร้างโมเดล

- ด้วยเหตุนี้ ตัวแปรระดับผู้สมัครทั้งหมด จึงไม่ย้ายศูนย์กลางไปที่ค่าเฉลี่ยกลุ่ม (Group-mean centering)
- การย้ายศูนย์กลาง จะย้ายเฉพาะตัวแปร IQ ที่ย้ายไปที่ค่า 100 (ค่าเฉลี่ยในประชากร) เท่านั้น ตัวแปรอื่นจะไม่มีการย้ายศูนย์กลาง

# โมเดลทดสอบสมมติฐานเป็นดังนี้



```
> out2b0 <- lmer(score ~ 1 + eesex + eesalary + eeworkexp + I(eeiq - 100)
+                + ersex + erexp + eesex:ersex
+                + (1 + eesex|erid), data=dat2, REML=FALSE)
> summary(out2b0)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: score ~ 1 + eesex + eesalary + eeworkexp + I(eeiq - 100) + ersex +
    erexp + eesex:ersex + (1 + eesex | erid)
   Data: dat2

     AIC      BIC   logLik deviance df.resid
 60259.1  60345.7 -30117.6  60235.1     9988

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.9117 -0.6471 -0.0151  0.6477  3.3815

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 erid     (Intercept) 18.402   4.290
          eesex        1.842   1.357    0.02
 Residual             18.585   4.311
Number of obs: 10000, groups:  erid, 1000

Fixed effects:
                Estimate Std. Error t value
(Intercept)    73.668030   0.221986 331.859
eesex          -0.530397   0.136242  -3.893
eesalary       -0.008172   0.054475  -0.150
eeworkexp       0.337878   0.014556  23.212
I(eeiq - 100)   0.028475   0.003065   9.291
ersex           0.221768   0.297455   0.746
erexp           0.011689   0.011099   1.053
eesex:ersex     0.957125   0.192642   4.968
```

ทดสอบอิทธิพลของเพศ โดยใส่ตัวแปรควบคุมเรียบร้อยแล้ว

ให้โมเดลนี้เป็นโมเดลเริ่มต้น แล้วใช้การอุปนัยเพิ่มเติมเพื่อหาความสัมพันธ์ระหว่างตัวแปรเพิ่มเติม

# ลดอิทธิพลตัวแปรอิสระทั้ง L1 และ L2

**Check Each Covariate**

**โมเดล 0**
L1 IV: เพศผู้สมัคร
เงินเดือนคาดหวัง
ประสบการณ์ทำงาน
เชาวน์ปัญญา
L2 IV: เพศผู้สัมภาษณ์
ประสบการณ์สัมภาษณ์
L1xL2: เพศผู้สมัคร x เพศผู้สัมภาษณ์
ชันสุ่ม: เพศผู้สมัคร

```
1 + eesex
+ eesalary
+ eeworkexp
+ I(eeiq - 100)
+ ersex
+ erexp
+ eesex:ersex
+ (1 + eesex|erid)
```

$p = .039$ → **โมเดล 1**: ชันสุ่ม: เงินเดือนคาดหวัง — Build-up strategy

$p = .085$ → **โมเดล 2**: ชันสุ่ม: ประสบการณ์ทำงาน — Build-up strategy

$p < .001$ → **โมเดล 3**: ~~L1 IV: ประสบการณ์ทำงาน~~ — Tear-down strategy

$p = .881$ → **โมเดล 4**: ชันสุ่ม: เชาวน์ปัญญา — Build-up strategy

$p < .001$ → **โมเดล 5**: ~~L1 IV: เชาวน์ปัญญา~~ — Tear-down strategy

$p = .292$ → **โมเดล 6**: ~~L2 IV: ประสบการณ์สัมภาษณ์~~ — Tear-down strategy

เงินเดือนคาดหวังมีความชันสุ่ม และนำประสบการณ์สัมภาษณ์ออกจากโมเดล

# ทดสอบปฏิสัมพันธ์ระหว่างชั้น

**Build-up Strategy**

**โมเดล 7**

L1 IV: เพศผู้สมัคร

เงินเดือนคาดหวัง

ประสบการณ์ทำงาน

เชาวน์ปัญญา

L2 IV: เพศผู้สัมภาษณ์

L1xL2: เพศผู้สมัคร x เพศผู้สัมภาษณ์

ชันสุ่ม: เพศผู้สมัคร

เงินเดือนคาดหวัง

```
1 + eesex
+ eesalary
+ eeworkexp
+ I(eeiq - 100)
+ ersex
+ eesex:ersex
+ (1 + eesex
+ eesalary|erid)
```

$p$ = .552

**โมเดล 10**

L1xL2: เงินเดือนคาดหวัง x เพศผู้สัมภาษณ์

```
1 + eesex
+ eesalary
+ eeworkexp
+ I(eeiq - 100)
+ ersex
+ eesex:ersex
+ eesalary:ersex
+ (1 + eesex
+ eesalary|erid)
```

# ทดสอบปฏิสัมพันธ์ระหว่างชั้น

**Build-up Strategy**

**โมเดล 9**

L1 IV: เพศผู้สมัคร

เงินเดือนคาดหวัง

ประสบการณ์ทำงาน

เชาวน์ปัญญา

L2 IV: เพศผู้สัมภาษณ์

ประสบการณ์สัมภาษณ์

L1xL2: เพศผู้สมัคร x เพศผู้สัมภาษณ์

ชั้นสุ่ม: เพศผู้สมัคร

เงินเดือนคาดหวัง

```
1 + eesex
+ eesalary
+ eeworkexp
+ I(eeiq - 100)
+ ersex
+ erexp
+ eesex:ersex
+ (1 + eesex
+ eesalary|erid)
```

$p$ = .412

**โมเดล 10**

L1xL2: เงินเดือนคาดหวัง x ประสบการณ์สัมภาษณ์

```
+ eesalary:erexp
```

$p$ = .470

**โมเดล 11**

L1xL2: เพศผู้สมัคร x ประสบการณ์สัมภาษณ์

```
+ eesex:erexp
```

```
> summary(out2m9)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: score ~ 1 + eesex + eeworkexp + I(eeiq - 100) + eesalary + ersex +
    eesex:ersex + (1 + eesex + eesalary | erid)
   Data: dat2
Control: lmerControl(optimizer = "Nelder_Mead")

     AIC      BIC   logLik deviance df.resid
 60255.9  60356.9 -30114.0  60227.9     9986

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8783 -0.6484 -0.0112  0.6418  3.4042

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 erid     (Intercept) 18.4288  4.2929
          eesex        1.8574  1.3628    0.01
          eesalary     0.2633  0.5131    0.05 -0.09
 Residual             18.3223  4.2805
Number of obs: 10000, groups:  erid, 1000

Fixed effects:
               Estimate Std. Error t value
(Intercept)   73.707032   0.219819 335.309
eesex         -0.532468   0.136162  -3.911
eeworkexp      0.337490   0.014568  23.166
I(eeiq - 100)  0.028619   0.003063   9.342
eesalary      -0.009272   0.057028  -0.163
ersex          0.222478   0.297553   0.748
eesex:ersex    0.956682   0.192611   4.967
```

โมเดลสุดท้าย หลังจากผ่านการค้นหาโมเดลด้วยวิธีอุปนัยเพิ่มเติมแล้ว

# ปัญหาจากตัวแปรอิสระที่ ICC = 0

- Snijders & Bosker (2012) ได้บอกว่า ให้หลีกเลี่ยง Group-mean centering ยกเว้นแต่จะมีทฤษฎีที่แสดงให้เห็นถึง Frog-pond effect เท่านั้น
- เหตุผลที่พวกเขาอ้าง คือ หากทำ Group-mean centering แล้ว อาจทำให้โมเดลคำนวณความแปรปรวนของค่าคงเหลือผิดพลาด
- ยกตัวอย่างการค่อยๆ สร้างโมเดล โดยทดลองใส่ตัวแปร Group-mean centering ลองสังเกตค่า $R_t^{2(m)}$ และ $\tau_{00}$

# ทดสอบตัวแปรอิสระระดับที่ 1

`consume ~ 1 + (1|groupid)`

**โมเดล 0**

IV: ไม่มี

$p < .001$

**โมเดล 1**

L1 IV: ความยาว (GMC)

`consume ~ 1 + difflengthx + (1|groupid)`

$$\tau_{00} = 24.04, \sigma^2 = 24.63$$

| $f_1$ | $f_2$ | $v$ | $m$ |
|---|---|---|---|
| .00 | .00 | .00 | .49 |

$$\tau_{00} = 27.49, \sigma^2 = 6.42$$

| $f_1$ | $f_2$ | $v$ | $m$ |
|---|---|---|---|
| .32 | .00 | .00 | .55 |

ตามหลักแล้ว `difflengthx` เป็นตัวแปรภายในกลุ่ม
จึงควรมีผลกระทบเฉพาะความแปรปรวนระดับที่ 1
ไม่ควรมีส่วนไปเพิ่มความแปรปรวนระดับที่ 2

# ปัญหาจากตัวแปรอิสระที่ ICC = 0

- ปรากฏการณ์นี้เกิดจากว่า การวิเคราะห์พหุระดับมีข้อตกลงเบื้องต้นว่า การสุ่มข้อมูลต้องเป็นแบบ 2 ระดับ (Two-stage Sampling)
  - สุ่มกลุ่มมาจากประชากรของกลุ่มก่อน
  - สุ่มบุคคลมาจากประชากรของแต่ละกลุ่ม
- ดังนั้น ตามข้อตกลงเบื้องต้นนี้ จะคาดหวังว่าค่าเฉลี่ยของตัวแปรอิสระภายในแต่ละกลุ่ม ต้องมีความแตกต่างกันบ้าง (ความแตกต่างจากการสุ่ม)
- การวิเคราะห์แบบความเป็นไปได้สูงสุด (Maximum Likelihood) จะมีการปรับค่าความแปรปรวนแต่ละระดับ เพื่อจัดการกับความแตกต่างจากการสุ่มที่เกิดขึ้น

# ปัญหาจากตัวแปรอิสระที่ ICC = 0

- แต่หากผู้วิจัยใส่ตัวแปรที่ไม่มีความแตกต่างระหว่างกลุ่มเลย เช่น
  - การจัดกระทำของผู้วิจัย เช่น บุคคลในกลุ่มครึ่งหนึ่งไปกลุ่มทดลอง และอีกครึ่งหนึ่งไปกลุ่มควบคุม และเป็นแบบนี้ทุกกลุ่ม
  - เวลาในการเก็บข้อมูล เช่น ทุกคนทุกวัดเวลา 6, 12, 18 เดือนเหมือนกันหมด
  - ตัวแปรที่ผ่านการย้ายศูนย์กลางไปที่ค่าเฉลี่ยกลุ่ม (Group-mean centered)

## ตัวแปรอิสระ ICC ในประชากร = 0

ตัวแปรอิสระมีค่าต่างกันระหว่างกลุ่มบ้าง จาก Random sampling

↓

อิทธิพลในกลุ่ม = อิทธิพลระหว่างกลุ่ม

↓

ค่าคงเหลือระดับกลุ่มลดลงบ้าง จากความแปรปรวนของ $X$ ที่เกิดจากการสุ่ม

↓

เมื่อเห็นว่ามันลดลงบ้าง ML จึงปรับค่าให้ $\tau_{00}$ สูงขึ้นเล็กน้อย (เทียบเท่ากับโมเดลเปล่า) เพื่อชดเชยความสัมพันธ์ลวงจาก $X$ ที่เกิดจากการสุ่ม

## ตัวแปรที่ทำ Group-mean centering

ตัวแปรอิสระต่างกันไม่แตกต่างกันเลย

↓

อิทธิพลในกลุ่ม = อิทธิพลระหว่างกลุ่ม

↓

ค่าคงเหลือระดับกลุ่มไม่ลดลงเลย

↓

ML ดันไปปรับค่าเหมือนกรณีทางซ้าย

↓

ปรับทั้งที่ไม่ควรปรับ ทำให้ค่าคงเหลือเพิ่มขึ้นจากที่ควรเป็น

# ปัญหาจากตัวแปรอิสระที่ ICC = 0

- จากเหตุการณ์ข้างบน Snijders & Bosker (2012) ถึงขนาดเสนอว่า พยายามหลีกเลี่ยง $R_w^2$ โดยให้ตีความหมาย $R_t^2$ แทน ซึ่งผมคิดว่าดูสถานการณ์เอา ไม่ใช่ยึดตามนี้ตายตัว
- เราต้องแยกปัญหานี้เป็น 2 แบบ
  1. ตัวแปรที่ ICC = 0 ตามธรรมชาติอยู่แล้ว เช่น การจัดกระทำกลุ่มทดลอง หรือ ตัวแปรเวลา
  2. ตัวแปรที่ไม่ได้มี ICC = 0 แต่บังคับให้ ICC = 0 เช่น การย้ายศูนย์กลางไปที่ค่าเฉลี่ยกลุ่ม ซึ่งสามารถแบ่งย่อยได้อีก 2 มุม คือ
     a) ค่าเฉลี่ยกลุ่มมีอิทธิพลต่อตัวแปรตาม
     b) ค่าเฉลี่ยกลุ่มไม่มีอิทธิพลต่อตัวแปรตาม

# ปัญหาจากตัวแปรอิสระที่ ICC = 0

- กรณีที่ 1 ตัวแปรที่ ICC = 0 ตามธรรมชาติอยู่แล้ว การเฟ้อของความแปรปรวนจะเป็นเรื่องปกติอยู่แล้ว แต่ค่าของการเฟ้อจะน้อยมาก ค่าของ $R_w^2$ จะมีผลกระทบน้อยมาก จนสามารถตีความได้ ไม่ต้องซีเรียส
- กรณีที่ 2b ตัวแปรที่ ICC ไม่เท่ากับ 0 แต่ใช้ Group-mean centering แล้วความเป็นจริง ค่าเฉลี่ยตัวแปรอิสระนั้นมีผลต่อตัวแปรตามจริง แต่ไม่ได้ใส่เข้าไปในโมเดล
  - ปัญหานี้เกิดจากการระบุโมเดลผิดพลาด (Model misspecification) ที่ไม่ใส่ตัวแปรที่สำคัญเข้าไปในโมเดล
  - แก้ไขโดยใส่ตัวแปรค่าเฉลี่ยของตัวแปรอิสระนั้นเข้าไปเป็นตัวแปรอิสระ
  - ในการค่อยๆ สร้างโมเดลข้างต้น สุดท้ายจะใส่ค่าเฉลี่ยกลุ่มอยู่ดี ดังนั้นใส่ตัวแปร Group-mean centering ก่อน แล้วใส่ค่าเฉลี่ยทีหลังได้

# ปัญหาจากตัวแปรอิสระที่ ICC = 0

- กรณีที่ 2a ตัวแปรที่ ICC ไม่เท่ากับ 0 แต่ใช้ Group-mean centering แล้วค่าเฉลี่ยของตัวแปรอิสระไม่มีผลต่อตัวแปรตาม
  - ต้องพิสูจน์ก่อนว่า ค่าเฉลี่ยของตัวแปรอิสระนั้นไม่มีผลอย่างมีนัยสำคัญต่อตัวแปรตาม (ใส่ก่อนแล้วค่อยเอาออกจากโมเดล)
  - กรณีนี้ ค่าของการเฟ้อจะน้อยมาก ค่าของ $R_w^2$ จะมีผลกระทบน้อยมาก จนสามารถตีความได้ ไม่ต้องซีเรียส
- ดังนั้น ไม่ต้องซีเรียสเรื่องปัญหาการเฟ้อของ $\tau_{00}$ หากใส่ตัวแปรให้ครบทุกตัว

# คาบต่อไป

- โมเดลตัวแปรแฝงพหุระดับ
- การบ้านที่ 7