การวิเคราะห์องค์ประกอบเชิงยืนยัน ตอนที่ 1 (Confirmatory Factor Analysis: Part 1)

สันทัด พรประเสริฐมานิต

โครงร่างการนำเสนอ

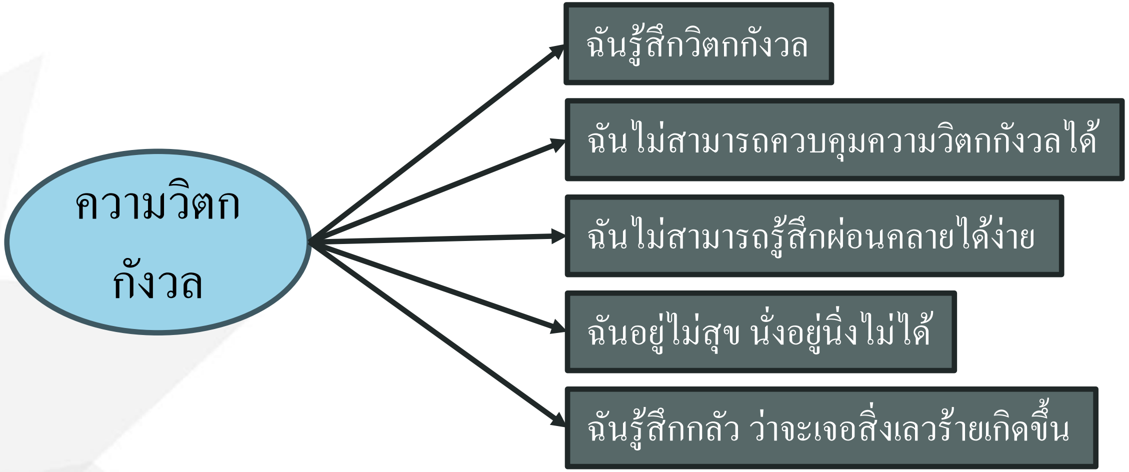
- แนะนำเนื้อหา
- ทฤษฎีการวัด
- การประมาณค่าจากความเป็นไปได้สูงสุด
- การกำหนดขนาดขององค์ประกอบ
- ค่าพารามิเตอร์ที่ถูกทำให้เป็นมาตรฐาน
- สัดส่วนร่วม
- การระบุโมเดล
- องค์ประกอบที่มีตัวบ่งชี้เดียว

แนะนำเนื้อหา

- องค์ประกอบ (Factor), ตัวแปรแฝง (Latent Variables), ภาวะสันนิษฐาน (Constructs) เป็นสิ่งที่ไม่สามารถวัดได้โดยตรง แต่เราใช้สิ่งที่เราสังเกตได้ อ้างอิงถึงองค์ประกอบเหล่านั้น
- ตัวแปรทางจิตวิทยา เช่น ความฉลาด ก็ไม่สามารถสังเกตได้โดยตรง แต่เราอ้างอิงจากพฤติกรรมที่สังเกตได้
- ตัวที่บ่งชี้ (Indicators) เช่น คะแนนแบบทดสอบ ผลการสังเกตพฤติกรรม คะแนนการตอบข้อสอบ เป็นสื่อกลางที่เราใช้อ้างอิงถึงองค์ประกอบเหล่านั้น

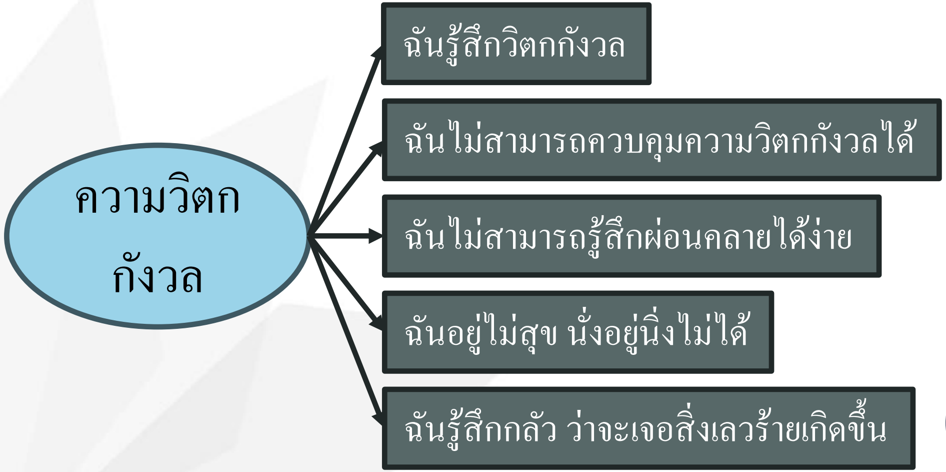
แนะนำเนื้อหา

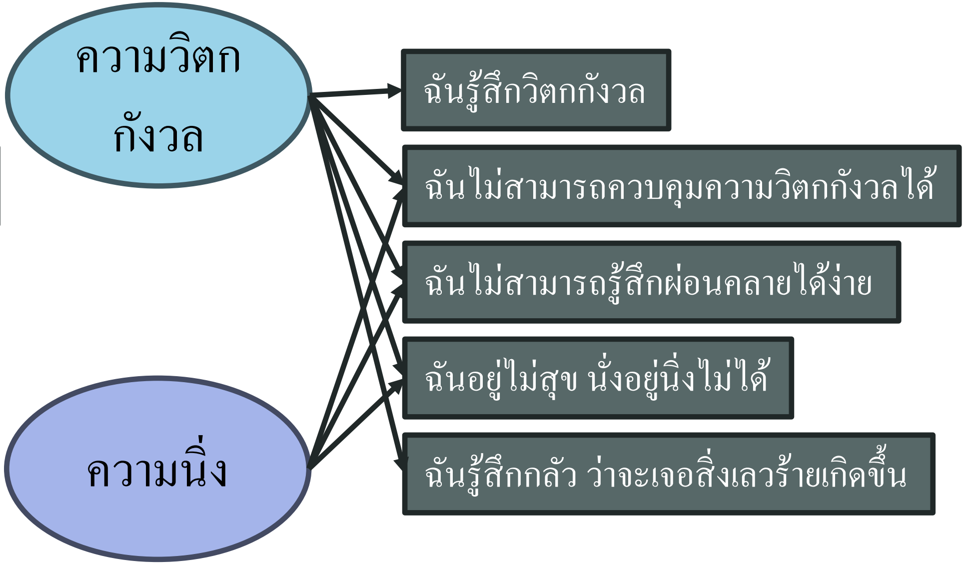
- การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) คือ การทดสอบสมมติฐานว่าความสัมพันธ์ระหว่างตัวบ่งชี้ (Indictors) และองค์ประกอบ (Factors) ที่ได้จากข้อมูลสอดคล้องกับสมมติฐานที่ตั้งไว้หรือไม่
  - เช่น ความวิตกกังวลเป็นองค์ประกอบที่สะท้อนการเปลี่ยนแปลงคะแนนของข้อคำถาม 5 ข้อ



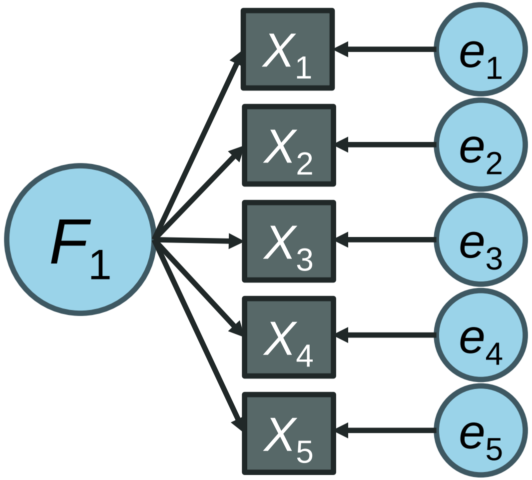
แนะนำเนื้อหา

- หลักการของ CFA คือ จะตรวจสอบว่า มีเพียงองค์ประกอบเดียวที่อธิบายตัวบ่งชี้ทั้ง 5 ข้อใช่หรือไม่ สมมติฐานนี้ ใกล้เคียงกับความสัมพันธ์ระหว่างตัวบ่งชี้ ที่เก็บข้อมูลมาหรือไม่
- มีองค์ประกอบรูปแบบอื่น ที่สามารถอธิบายความสัมพันธ์ระหว่างตัวบ่งชี้เหล่านี้หรือไม่





เครื่องหมาย =~ แสดงว่าองค์ประกอบอะไร ถูกอธิบายด้วยตัวบ่งชี้อะไร




```
> library(lavaan)
>
> dat <- read.table("lecture8ex1.csv", sep=",", header=TRUE)
>
> m1 <- '
+ f1 =~ x1 + x2 + x3 + x4 + x5
+ '
```

ใช้คำสั่ง cfa เพื่อวิเคราะห์ CFA

```
> out <- cfa(m1, data=dat)
> summary(out)
lavaan 0.6.15 ended normally after 26 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        10

  Number of observations                           200

Model Test User Model:

  Test statistic                                21.923
  Degrees of freedom                                 5
  P-value (Chi-square)                           0.001

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured
```

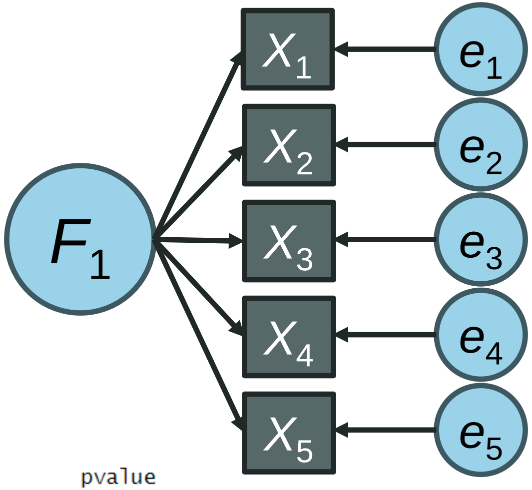
```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    x1                1.000
    x2                1.916    0.443    4.322    0.000
    x3                2.198    0.508    4.324    0.000
    x4                1.438    0.355    4.048    0.000
    x5                0.833    0.265    3.138    0.002

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.409    0.043    9.428    0.000
   .x2                0.228    0.039    5.853    0.000
   .x3                0.290    0.051    5.727    0.000
   .x4                0.319    0.038    8.437    0.000
   .x5                0.382    0.040    9.578    0.000
    f1                0.066    0.028    2.337    0.019
```

น้ำหนักองค์ประกอบ

ความแปรปรวนของ e_1-e_5

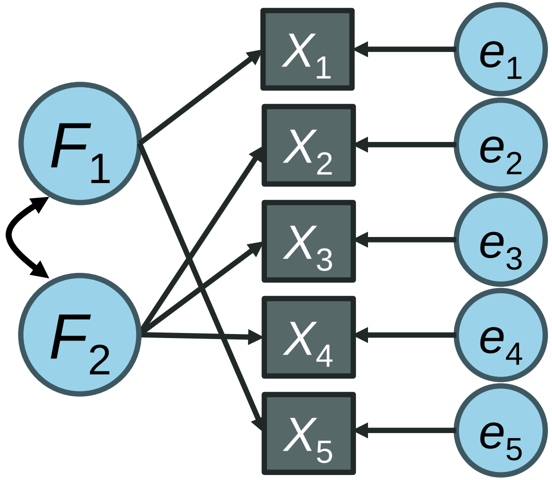
ความแปรปรวนของ F_1

โมเดลที่มีองค์ประกอบเดียว ดูแล้วไม่สามารถ
อธิบายความสัมพันธ์ระหว่างตัวแปรจากข้อมูล
ที่เก็บมาได้ดี จากดัชนีความเหมาะสม




```
fitmeasures(out)
               npar                fmin               chisq                  df              pvalue
             10.000               0.055              21.923               5.000               0.001
     baseline.chisq         baseline.df   baseline.pvalue                   cfi                 tli
            167.092              10.000               0.000               0.892               0.785
               nnfi                 rfi                 nfi                pnfi                 ifi
              0.785               0.738               0.869               0.434               0.896
                rni                logl   unrestricted.logl                 aic                 bic
              0.892            -982.474            -971.513            1984.949            2017.932
             ntotal                bic2               rmsea      rmsea.ci.lower      rmsea.ci.upper
            200.000            1986.251               0.130               0.078               0.188
     rmsea.ci.level        rmsea.pvalue      rmsea.close.h0 rmsea.notclose.pvalue  rmsea.notclose.h0
              0.900               0.008               0.050               0.942               0.080
                rmr          rmr_nomean                srmr        srmr_bentler  srmr_bentler_nomean
              0.029               0.029               0.061               0.061               0.061
               crmr         crmr_nomean          srmr_mplus   srmr_mplus_nomean               cn_05
              0.075               0.075               0.061               0.061             101.995
              cn_01                 gfi                agfi                pgfi                 mfi
            138.630               0.960               0.880               0.320               0.959
               ecvi
              0.210
```

```
> m2 <- '
+ f1 =~ x1 + x5
+ f2 =~ x2 + x3 + x4
+ '
> out2 <- cfa(m2, data=dat)
```




```
> out2 <- cfa(m2, data=dat)
> summary(out2)
lavaan 0.6.15 ended normally after 26 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        11

  Number of observations                           200

Model Test User Model:

  Test statistic                                19.550
  Degrees of freedom                                 4
  P-value (Chi-square)                           0.001

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured
```

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    x1                1.000
    x5                0.824    0.261    3.153    0.002
  f2 =~
    x2                1.000
    x3                1.126    0.171    6.576    0.000
    x4                0.740    0.124    5.969    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 ~~
    f2                0.122    0.031    3.899    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.359    0.055    6.496    0.000
   .x5                0.349    0.045    7.749    0.000
   .x2                0.221    0.040    5.546    0.000
   .x3                0.293    0.051    5.706    0.000
   .x4                0.319    0.038    8.430    0.000
    f1                0.116    0.052    2.214    0.027
    f2                0.248    0.053    4.677    0.000
```

น้ำหนักองค์ประกอบ

ความแปรปรวนร่วม

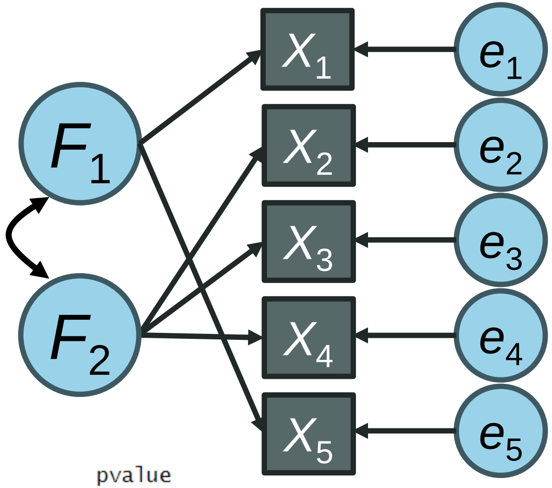
ระหว่างองค์ประกอบ

ความแปรปรวนของ

e_1-e_5

ความแปรปรวนของ F_1, F_2

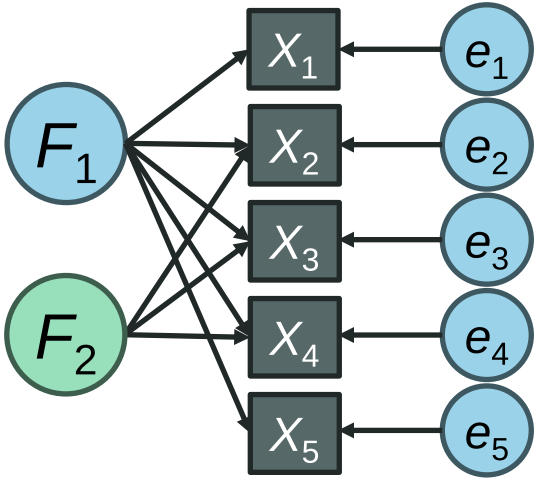
โมเดลที่มีสององค์ประกอบ ดูแล้วยังคงไม่สามารถ
อธิบายความสัมพันธ์ระหว่างตัวแปรจากข้อมูล
ที่เก็บมาได้ เพราะดัชนีความเหมาะสมไม่ดี




```
> fitmeasures(out2)
               npar                fmin               chisq                  df              pvalue 
             11.000               0.049              19.550               4.000               0.001 
     baseline.chisq         baseline.df     baseline.pvalue                 cfi                 tli 
            167.092              10.000               0.000               0.901               0.753 
               nnfi                 rfi                 nfi                pnfi                 ifi 
              0.753               0.707               0.883               0.353               0.905 
                rni                logl   unrestricted.logl                 aic                 bic 
              0.901            -981.288            -971.513            1984.576            2020.858 
             ntotal                bic2               rmsea      rmsea.ci.lower      rmsea.ci.upper 
            200.000            1986.009               0.139               0.082               0.204 
     rmsea.ci.level        rmsea.pvalue      rmsea.close.h0 rmsea.notclose.pvalue  rmsea.notclose.h0 
              0.900               0.008               0.050               0.954               0.080 
                rmr          rmr_nomean                srmr         srmr_bentler  srmr_bentler_nomean 
              0.027               0.027               0.056               0.056               0.056 
               crmr         crmr_nomean          srmr_mplus   srmr_mplus_nomean               cn_05 
              0.069               0.069               0.056               0.056              98.060 
              cn_01                 gfi                agfi                pgfi                 mfi 
            136.821               0.962               0.859               0.257               0.962 
               ecvi 
              0.208 
```

กำหนดให้ความแปรปรวนร่วม (ความสัมพันธ์) ระหว่างองค์ประกอบเป็น 0

```
> m3 <- '
+ f1 =~ x1 + x2 + x3 + x4 + x5
+ f2 =~ x2 + x3 + x4
+ f1 ~~ 0*f2
+ '
> out3 <- cfa(m3, data=dat)
```

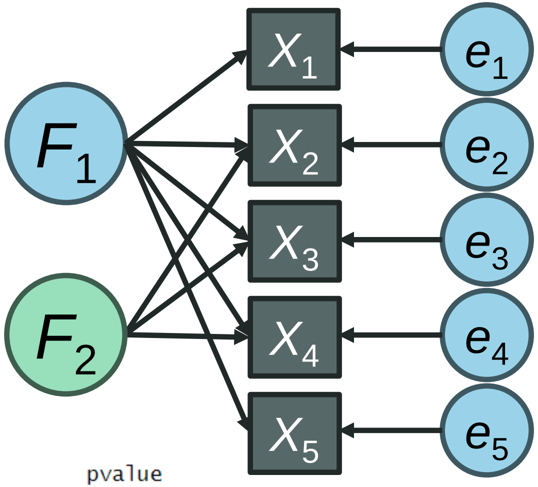



```
> summary(out3)
lavaan 0.6.15 ended normally after 40 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        13

  Number of observations                           200

Model Test User Model:

  Test statistic                                13.817
  Degrees of freedom                                 2
  P-value (Chi-square)                           0.001

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured
```

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    x1                1.000
    x2                0.834    0.343    2.431    0.015
    x3                1.380    0.491    2.809    0.005
    x4                0.840    0.342    2.457    0.014
    x5                0.809    0.242    3.344    0.001
  f2 =~
    x2                1.000
    x3                0.405    0.387    1.046    0.296
    x4                0.324    0.312    1.039    0.299

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 ~~
    f2                0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.357    0.054    6.573    0.000
   .x2                0.027    0.286    0.094    0.925
   .x3                0.324    0.057    5.651    0.000
   .x4                0.333    0.040    8.354    0.000
   .x5                0.351    0.044    7.954    0.000
    f1                0.118    0.051    2.288    0.022
    f2                0.360    0.281    1.281    0.200
```

ความแปรปรวนร่วมระหว่างองค์ประกอบถูกกำหนดให้คงที่ มีค่าเท่ากับ 0

โมเดลที่มีสององค์ประกอบแบบที่มีองค์ประกอบรวม
และมีองค์ประกอบย่อย ที่ไม่สัมพันธ์กับองค์ประกอบรวม
ซึ่งจะเรียกว่าโมเดล bifactor ดูแล้วก็ยังไม่สามารถ
อธิบายความสัมพันธ์ระหว่างตัวแปรจากข้อมูลที่เก็บมาได้ดี




```
> fitmeasures(out3)
           npar             fmin             chisq                 df              pvalue
         13.000            0.035            13.817              2.000               0.001
 baseline.chisq      baseline.df  baseline.pvalue                cfi                 tli
        167.092           10.000             0.000              0.925               0.624
           nnfi              rfi               nfi               pnfi                 ifi
          0.624            0.587             0.917              0.183               0.928
            rni             logl  unrestricted.logl               aic                 bic
          0.925         -978.422          -971.513           1982.843            2025.721
         ntotal             bic2             rmsea     rmsea.ci.lower      rmsea.ci.upper
        200.000         1984.536             0.172              0.094               0.262
 rmsea.ci.level     rmsea.pvalue    rmsea.close.h0 rmsea.notclose.pvalue  rmsea.notclose.h0
          0.900            0.007             0.050              0.972               0.080
            rmr       rmr_nomean              srmr       srmr_bentler   srmr_bentler_nomean
          0.023            0.023             0.049              0.049               0.049
           crmr      crmr_nomean        srmr_mplus  srmr_mplus_nomean               cn_05
          0.060            0.060             0.049              0.049              87.724
          cn_01              gfi              agfi               pgfi                 mfi
        134.315            0.974             0.805              0.130               0.971
           ecvi
          0.199
```

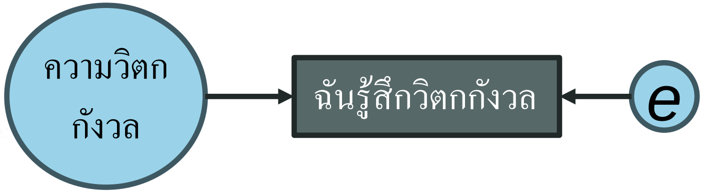
แนะนำเนื้อหา

- โมเดลที่อยู่ภายใต้ 5 ข้อคำถามนี้ ยังไม่ได้ข้อสรุปว่าโมเดลเป็นโมเดลที่เหมาะสมกับโมเดล เพราะข้อมูลความเหมาะสมของโมเดลแต่ละโมเดลต่อข้อมูลยังไม่ดีพอ
- เป็นเรื่องธรรมดา ที่การวิเคราะห์โมเดลจะได้ผลออกมาว่าไม่เจอโมเดลใดเหมาะสมเลย นักวิจัยอาจเลือกที่จะหาโมเดลใหม่ โดยอาจปรับข้อมูลจากโมเดลเดิม หรือคัดเลือกโมเดลที่ดีที่สุดตามที่มี
- โดยทั่วไป การทำ CFA มีเพียงแค่กำหนดโมเดล แล้วทดสอบว่าโมเดลเหมาะสมกับข้อมูลที่ได้มาหรือไม่เท่านั้น

ทฤษฎีการวัด

- การตอบคำถามหนึ่ง การแสดงสิ่งเร้าหนึ่ง เกิดจากปัจจัยหลายๆ ปัจจัย เช่น การตอบคำถาม “ฉันรู้สึกวิตกกังวล” ที่ต้องตอบว่าเห็นด้วยหรือไม่เห็นด้วย ขึ้นอยู่กับหลายปัจจัย
- บุคลิกภาพวิตกกังวล เป็นปัจจัยหนึ่ง แต่ยังมีปัจจัยอื่นอีกมากมาย เช่น อารมณ์ตอนปัจจุบัน อุณหภูมิห้อง ความแตกฉานทางภาษา ฯลฯ

คะแนนที่ตอบ = องค์ประกอบที่ 1 + 2 + 3 + ... = องค์ประกอบที่สนใจ + องค์ประกอบที่ไม่สนใจ (error)



เราคาดหวังให้องค์ประกอบที่สนใจ สามารถอธิบายการเปลี่ยนแปลงของคะแนนที่ตอบได้อย่างมีนัยสำคัญ และจะยิ่งดี หากองค์ประกอบที่สนใจ อธิบายการเปลี่ยนแปลงของคะแนนที่เก็บข้อมูลมาเป็นส่วนใหญ่

ทฤษฎีการวัด

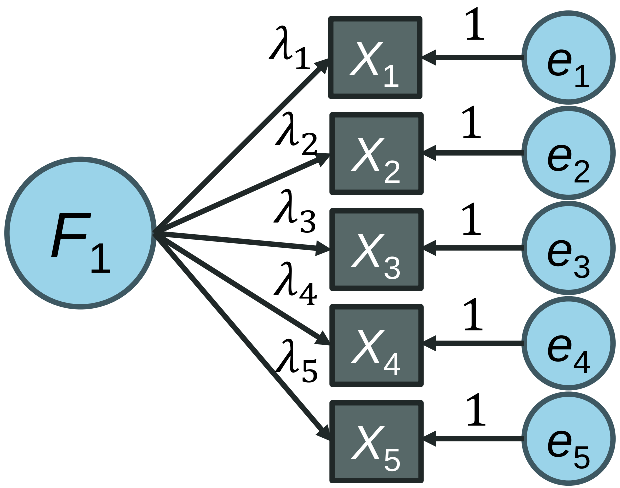
- ในทฤษฎีการวัด (Test Theory) จะกล่าวว่า คะแนนที่สังเกตได้ (Observed Score; *O*) เกิดจากคะแนนที่แท้จริง (True Score; *T*) รวมกับความผิดพลาด (Error; *E*)

$$\boxed{O = T + E}$$

- ทั้ง *T* และ *E* ล้วนเป็นองค์ประกอบทั้งหมด องค์ประกอบที่เราสนใจ เราจะจัดให้เป็น *T* ส่วนองค์ประกอบที่ไม่สนใจจะจัดให้เป็น *E*
- ในแผนภาพ วงกลม จะหมายถึงองค์ประกอบ หรือสิ่งที่ไม่สามารถวัดหรือสังเกตได้โดยตรง ส่วนสี่เหลี่ยม จะหมายถึง ตัวบ่งชี้ หรือสิ่งที่สามารถวัดได้โดยตรง

$$O_{i1} = T_i + e_{i1}$$
$$O_{i2} = T_i + e_{i2}$$
$$O_{i3} = T_i + e_{i3}$$
$$O_{i4} = T_i + e_{i4}$$
$$O_{i5} = T_i + e_{i5}$$

$$X_{i1} = \lambda_1 F_i + e_{i1}$$
$$X_{i2} = \lambda_2 F_i + e_{i2}$$
$$X_{i3} = \lambda_3 F_i + e_{i3}$$
$$X_{i4} = \lambda_4 F_i + e_{i4}$$
$$X_{i5} = \lambda_5 F_i + e_{i5}$$



สมการนี้เป็นเพียงรูปแบบหนึ่งของการอธิบาย $O = T + E$

O_{ij} = คะแนนที่สังเกตได้ ของตัวแปรที่ j คนที่ i

e_{ij} = คะแนนความผิดพลาด ของตัวแปรที่ j คนที่ i

T_i = คะแนนที่แท้จริง (คะแนนองค์ประกอบ) ของคนที่ i

X_{ij} = คำตอบของข้อคำถามที่ j คนที่ i

F_i = คะแนนองค์ประกอบ ของคนที่ i

λ_j = น้ำหนักองค์ประกอบของข้อที่ j

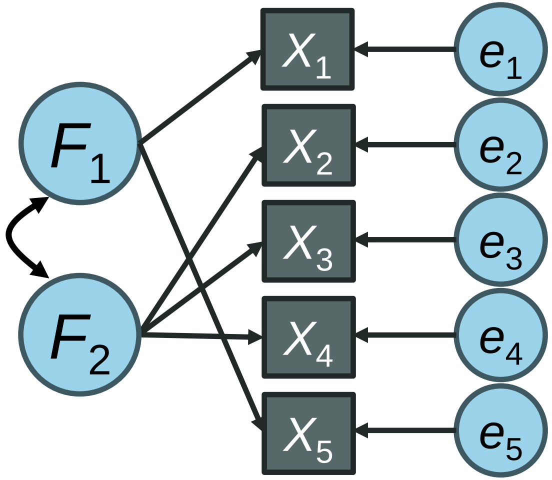
$$X_{i1} = \lambda_{11}F_{i1} + 0 \cdot F_{i2} + e_{i1} = \lambda_{11}F_{i1} + e_{i1}$$

$$X_{i2} = 0 \cdot F_{i1} + \lambda_{22}F_{i2} + e_{i2} = \lambda_{22}F_{i2} + e_{i2}$$

$$X_{i3} = 0 \cdot F_{i1} + \lambda_{32}F_{i2} + e_{i3} = \lambda_{32}F_{i2} + e_{i3}$$

$$X_{i4} = 0 \cdot F_{i1} + \lambda_{42}F_{i2} + e_{i4} = \lambda_{42}F_{i2} + e_{i4}$$

$$X_{i5} = \lambda_{51}F_{i1} + 0 \cdot F_{i2} + e_{i5} = \lambda_{51}F_{i1} + e_{i5}$$



โมเดลนี้ให้องค์ประกอบมีความสัมพันธ์กัน

F_{ik} = คะแนนองค์ประกอบที่ k ของคนที่ i

λ_{jk} = น้ำหนักองค์ประกอบของข้อที่ j จากองค์ประกอบที่ k

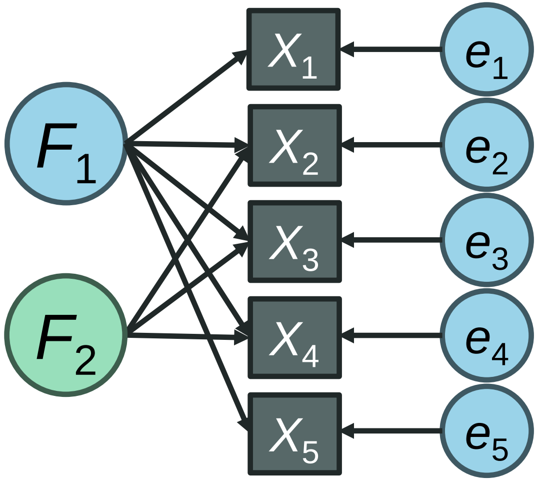
$$X_{i1} = \lambda_{11}F_{i1} + 0 \cdot F_{i2} + e_{i1} = \lambda_{11}F_{i1} + e_{i1}$$

$$X_{i2} = \lambda_{21}F_{i1} + \lambda_{22}F_{i2} + e_{i2} = \lambda_{21}F_{i1} + \lambda_{22}F_{i2} + e_{i2}$$

$$X_{i3} = \lambda_{31}F_{i1} + \lambda_{32}F_{i2} + e_{i3} = \lambda_{31}F_{i1} + \lambda_{32}F_{i2} + e_{i3}$$

$$X_{i4} = \lambda_{41}F_{i1} + \lambda_{42}F_{i2} + e_{i4} = \lambda_{41}F_{i1} + \lambda_{42}F_{i2} + e_{i4}$$

$$X_{i5} = \lambda_{51}F_{i1} + 0 \cdot F_{i2} + e_{i5} = \lambda_{51}F_{i1} + e_{i5}$$



โมเดลนี้กำหนด (Fix) องค์ประกอบไม่มีความสัมพันธ์กัน

F_{ik} = คะแนนองค์ประกอบที่ k ของคนที่ i

λ_{jk} = น้ำหนักองค์ประกอบของข้อที่ j จากองค์ประกอบที่ k

ทฤษฎีการวัด

- ปัญหาของการเข้าใจสมการคณิตศาสตร์ คือ ตัวห้อย (subscript) มีเยอะมาก จริงๆ แล้วมันมาจากการเรียนระบบสมการเชิงเส้น ซึ่งสามารถเขียนในรูปแบบเมทริกซ์ ตำแหน่งของแถวและคอลัมน์เป็นตัวบอกความหมายของตัวห้อย

$$X_{i1} = \lambda_1 F_i + e_{i1}$$

$$X_{i2} = \lambda_2 F_i + e_{i2}$$

$$X_{i3} = \lambda_3 F_i + e_{i3}$$

$$X_{i4} = \lambda_4 F_i + e_{i4}$$

$$X_{i5} = \lambda_5 F_i + e_{i5}$$

$$\begin{bmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \\ X_{i4} \\ X_{i5} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix} [F_i] + \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \\ e_{i5} \end{bmatrix}$$

$$\mathbf{x}_i = \mathbf{\Lambda f}_i + \mathbf{e}_i$$

$X_{i1} = \lambda_{11}F_{i1} + e_{i1}$

$X_{i2} = \lambda_{21}F_{i1} + \lambda_{22}F_{i2} + e_{i2}$

$X_{i3} = \lambda_{31}F_{i1} + \lambda_{32}F_{i2} + e_{i3}$

$X_{i4} = \lambda_{41}F_{i1} + \lambda_{42}F_{i2} + e_{i4}$

$X_{i5} = \lambda_{51}F_{i1} + e_{i5}$

$$\begin{bmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \\ X_{i4} \\ X_{i5} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & 0 \end{bmatrix} \begin{bmatrix} F_{i1} \\ F_{i2} \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \\ e_{i5} \end{bmatrix}$$

$$\mathbf{x}_i = \mathbf{\Lambda}\mathbf{f}_i + \mathbf{e}_i$$

รู้ค่า

ไม่รู้ค่า

$$\mathbf{x}_{p\times 1} = \begin{bmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \\ X_{i4} \\ X_{i5} \end{bmatrix} \qquad \underbrace{\mathbf{f}_{k\times 1} = \begin{bmatrix} F_{i1} \\ F_{i2} \end{bmatrix} \qquad \mathbf{\Lambda}_{p\times k} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & 0 \end{bmatrix} \qquad \mathbf{e}_{p\times 1} = \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \\ e_{i5} \end{bmatrix}}_{\text{ไม่รู้ค่า}}$$

จะหาค่าใน **f**, **Λ**, และ **e** อย่างไร เมื่อมีค่าที่เรารู้ค่าเพียงแค่ค่าของ **x**

การประมาณค่าจากความเป็นไปได้สูงสุด

- ปัญหาที่ไม่สามารถหาค่า **f**, $\boldsymbol{\Lambda}$, และ **e** ได้ จะเรียกว่า ความกำหนดไม่ได้ขององค์ประกอบ (Factor Indeterminacy) วิธีการแก้ไขของนักสถิติ คือ การเปลี่ยนจากการคำนวณจากข้อมูลดิบ เป็นการคำนวณจากเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix) ของตัวบ่งชี้ ดังนี้

$$\mathbf{x}_i = \boldsymbol{\Lambda}\mathbf{f}_i + \mathbf{e}_i$$

$$\mathrm{Cov}(\mathbf{x}) = \mathrm{Cov}(\boldsymbol{\Lambda}\mathbf{f} + \mathbf{e})$$

$$\mathrm{Cov}(\mathbf{x}) = \mathrm{Cov}(\boldsymbol{\Lambda}\mathbf{f}) + \mathrm{Cov}(\mathbf{e})$$

$$\mathrm{Cov}(\mathbf{x}) = (\boldsymbol{\Lambda}')'\mathrm{Cov}(\mathbf{f})\boldsymbol{\Lambda}' + \mathrm{Cov}(\mathbf{e})$$

$$\boldsymbol{\Sigma}_{p\times p} = \boldsymbol{\Lambda}_{p\times k}\boldsymbol{\Phi}_{k\times k}\boldsymbol{\Lambda}'_{p\times k} + \boldsymbol{\Theta}_{p\times p}$$

$\boldsymbol{\Phi}$ เป็นความแปรปรวนร่วมระหว่างองค์ประกอบ (**f**)

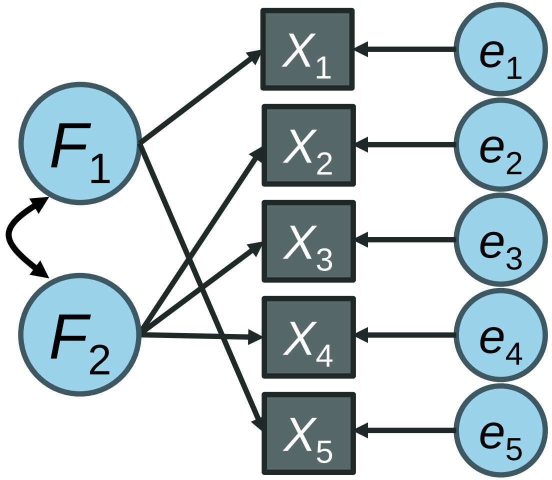
$\boldsymbol{\Theta}$ เป็นความแปรปรวนร่วมระหว่างค่าคงเหลือ (**e**)

$\boldsymbol{\Sigma}$ คือเมทริกซ์ความแปรปรวนร่วมที่ประมาณค่าจากโมเดล (Model-implied Covariance Matrix)

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{11} & 0 \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \\ \lambda_{51} & 0 \end{bmatrix} \qquad \boldsymbol{\Phi} = \begin{bmatrix} 1 & \\ \phi_{21} & 1 \end{bmatrix}$$

$$\boldsymbol{\Theta} = \begin{bmatrix} \theta_{11} & & & & \\ 0 & \theta_{22} & & & \\ 0 & 0 & \theta_{33} & & \\ 0 & 0 & 0 & \theta_{44} & \\ 0 & 0 & 0 & 0 & \theta_{55} \end{bmatrix}$$



$$\boldsymbol{\Sigma} = \begin{bmatrix} \lambda_{11}^2 + \theta_{11} & & & & \\ \lambda_{11}\phi_{21}\lambda_{22} & \lambda_{22}^2 + \theta_{22} & & & \\ \lambda_{11}\phi_{21}\lambda_{32} & \lambda_{22}\lambda_{32} & \lambda_{32}^2 + \theta_{33} & & \\ \lambda_{11}\phi_{21}\lambda_{42} & \lambda_{22}\lambda_{42} & \lambda_{32}\lambda_{42} & \lambda_{42}^2 + \theta_{44} & \\ \lambda_{11}\lambda_{51} & \lambda_{51}\phi_{21}\lambda_{22} & \lambda_{51}\phi_{21}\lambda_{32} & \lambda_{51}\phi_{21}\lambda_{42} & \lambda_{51}^2 + \theta_{55} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & & & & \\ \sigma_{21} & \sigma_{22} & & & \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{bmatrix}$$

การประมาณค่าจากความเป็นไปได้สูงสุด

ค่าพารามิเตอร์ที่ต้องประมาณค่า 11 ตัว

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & 0 \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \\ \lambda_{51} & 0 \end{bmatrix} \quad \mathbf{\Phi} = \begin{bmatrix} 1 & \\ \phi_{21} & 1 \end{bmatrix} \quad \mathbf{\Theta} = \begin{bmatrix} \theta_{11} & & & & \\ 0 & \theta_{22} & & & \\ 0 & 0 & \theta_{33} & & \\ 0 & 0 & 0 & \theta_{44} & \\ 0 & 0 & 0 & 0 & \theta_{55} \end{bmatrix}$$

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & & & & \\ \sigma_{21} & \sigma_{22} & & & \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{bmatrix}$$

นำพารามิเตอร์ทั้ง 11 ตัวไปสร้าง $\mathbf{\Sigma}$ เพื่อให้ใกล้เคียงกับข้อมูลมากที่สุด ด้วยวิธี Maximum Likelihood (ML)

การประมาณค่าจากความเป็นไปได้สูงสุด

- ML จะหาค่าพารามิเตอร์ทั้งหมด 11 ตัว ที่ทำให้มีค่า log-likelihood สูงที่สุด กล่าวคือ จะสร้าง Model-implied means ($\boldsymbol{\mu}$) (ถ้ามี) และ Model-implied covariance matrix ($\boldsymbol{\Sigma}$) ให้ใกล้เคียงกับข้อมูลมากที่สุด โดยใช้สมการ

$$\log L = -\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|\boldsymbol{\Sigma}| - \frac{N}{2}tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \frac{N}{2}[\boldsymbol{m}_y - \boldsymbol{\mu}]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{m}_y - \boldsymbol{\mu}]$$

- โมเดลนี้ประมาณค่า Covariance matrix อย่างเดียว ไม่ได้สร้างโมเดลไปจำกัดการประมาณค่าเฉลี่ย จะทำให้ Log-likelihood มีค่าดังนี้

$$\log L = -\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|\boldsymbol{\Sigma}| - \frac{N}{2}tr(\mathbf{S}\boldsymbol{\Sigma}^{-1})$$

- อย่าลืมว่าค่า log-likelihood จากโมเดลเป้าหมายจะต่ำกว่าโมเดลอิ่มตัวเสมอ

การประมาณค่าจากความเป็นไปได้สูงสุด

- F_{ML} จะเป็นฟังก์ชั่นจาก log-likelihood จากโมเดลอิ่มตัว และ log-likelihood จากโมเดลเป้าหมาย บางครั้งอาจเรียกว่าสมการค่าเบี่ยงเบน (Discrepancy Function) มีค่าดังนี้

$$F_{ML} = \frac{2}{N}(\log L_S - \log L_0) = \log|\boldsymbol{\Sigma}| - \log|\boldsymbol{S}| - p + tr(\mathbf{S}\boldsymbol{\Sigma}^{-1})$$

- การเขียนโปรแกรม จะหาค่าเป็นไปได้สูงสุดของ log-likelihood ของโมเดล หรือค่า F_{ML} ให้ต่ำที่สุดก็ได้ จะได้ค่าพารามิเตอร์เดียวกัน
- F_{ML} จะบ่งบอกถึงความเหมาะสมของโมเดลด้วย เพราะ เป็นการเปรียบเทียบระหว่างโมเดลที่สร้างขึ้นมา และโมเดลอิ่มตัว (Saturated Model) ที่ประมาณค่าเฉลี่ย และความแปรปรวนร่วมทั้งหมด

การประมาณค่าจากความเป็นไปได้สูงสุด

$$F_{ML} = tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) + \underbrace{\ln|\boldsymbol{\Sigma}| - \ln|\mathbf{S}|} - p$$

ถ้า $\mathbf{S} = \boldsymbol{\Sigma}$ แล้ว ค่านี้จะเท่ากับ $\mathbf{I}$ และผลรวมของสมาชิกแนวทแยง (trace) จะเท่ากับจำนวนข้อคำถาม แต่หากไม่เท่ากัน แตกต่างกันมาก ค่านี้จะยิ่งมีค่าสูง

ค่า Determinant ของเมทริกซ์ความแปรปรวนร่วมจะหมายถึง Generalized variance ของชุดตัวแปร ถ้า $\mathbf{S} = \boldsymbol{\Sigma}$ ค่าส่วนต่างนี้จะเท่ากับ 0 แต่ถ้าไม่เท่ากัน มีโอกาสเป็นทั้งบวกและลบ (ดูพฤติกรรมของค่านี้ได้ในสไลด์ถัดไป)

- วัตถุประสงค์ คือ ใช้คอมพิวเตอร์ หาค่าของสมาชิกใน $\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Theta}$ ไปเรื่อยๆ จนกว่าจะได้ค่า F_{ML} ที่ต่ำที่สุด (หรือหา -F_{ML} ที่สูงที่สุด)
- ดูเพิ่มเติมที่ https://www.quantpsy.org/misc/discrepancy_derivation_022316.pdf

การประมาณค่าจากความเป็นไปได้สูงสุด

- ทดสอบการคำนวณ F_{ML}

$$\mathbf{S} = \begin{bmatrix} 1 & 0.2 & 0.1 \\ 0.2 & 1 & 0.3 \\ 0.1 & 0.3 & 1 \end{bmatrix}$$

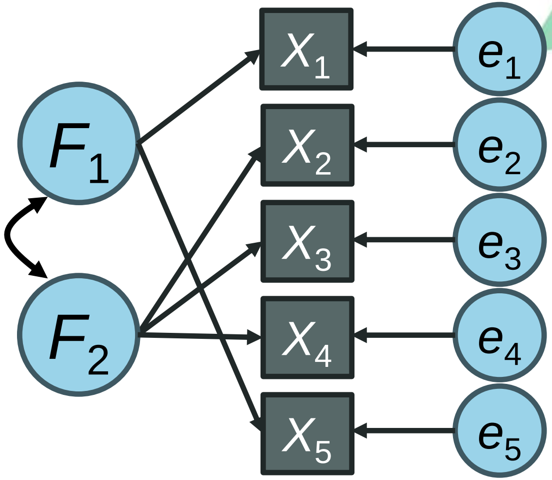
```r
> S <- matrix(c(1, 0.2, 0.1,
+               0.2, 1, 0.3,
+               0.1, 0.3, 1), nrow=3)
> Sigma1 <- Sigma2 <- Sigma3 <- S
> Sigma1[1,2] <- Sigma1[2,1] <- 0
> Sigma3[1,2] <- Sigma3[2,1] <- 0.4
> c(sum(diag(S %*% solve(Sigma1))),
+   sum(diag(S %*% solve(Sigma2))),
+   sum(diag(S %*% solve(Sigma3))))
[1] 3.013333 3.000000 3.193717
> c(log(det(Sigma1)) - log(det(S)),
+   log(det(Sigma2)) - log(det(S)),
+   log(det(Sigma3)) - log(det(S)))
[1]  0.03160534  0.00000000 -0.13222163
> c(sum(diag(S %*% solve(Sigma1))) + log(det(Sigma1)) - log(det(S)),
+   sum(diag(S %*% solve(Sigma2))) + log(det(Sigma2)) - log(det(S)),
+   sum(diag(S %*% solve(Sigma3))) + log(det(Sigma3)) - log(det(S)))
[1] 3.044939 3.000000 3.061496
```

$$\mathbf{\Sigma_1} = \begin{bmatrix} 1 & 0 & 0.1 \\ 0 & 1 & 0.3 \\ 0.1 & 0.3 & 1 \end{bmatrix} \quad \mathbf{\Sigma_2} = \begin{bmatrix} 1 & 0.2 & 0.1 \\ 0.2 & 1 & 0.3 \\ 0.1 & 0.3 & 1 \end{bmatrix} \quad \mathbf{\Sigma_3} = \begin{bmatrix} 1 & 0.4 & 0.1 \\ 0.4 & 1 & 0.3 \\ 0.1 & 0.3 & 1 \end{bmatrix}$$

$$tr\left(\mathbf{S\Sigma}_1^{-1}\right) = 3.013 \quad tr\left(\mathbf{S\Sigma}_2^{-1}\right) = 3.000 \quad tr\left(\mathbf{S\Sigma}_3^{-1}\right) = 3.194$$

$$\log\frac{|\mathbf{\Sigma_1}|}{|\mathbf{S}|} = 0.032 \quad \log\frac{|\mathbf{\Sigma_2}|}{|\mathbf{S}|} = 0.000 \quad \log\frac{|\mathbf{\Sigma_3}|}{|\mathbf{S}|} = -0.132$$

$$F_{ML(1)} = 0.045 \quad F_{ML(2)} = 0.000 \quad F_{ML(3)} = 0.061$$



$$\mathbf{\Phi} = \begin{bmatrix} 1 & \\ 0.72 & 1 \end{bmatrix}$$

$$\mathbf{\Lambda} = \begin{bmatrix} 0.35 & 0 \\ 0 & 0.50 \\ 0 & 0.56 \\ 0 & 0.36 \\ 0.28 & 0 \end{bmatrix}$$

$$\mathbf{\Theta} = \begin{bmatrix} 0.36 & & & & \\ 0 & 0.22 & & & \\ 0 & 0 & 0.29 & & \\ 0 & 0 & 0 & 0.32 & \\ 0 & 0 & 0 & 0 & 0.35 \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} 0.48 & & & & \\ 0.12 & 0.47 & & & \\ 0.14 & 0.28 & 0.61 & & \\ 0.09 & 0.18 & 0.21 & 0.45 & \\ 0.10 & 0.10 & 0.11 & 0.07 & 0.43 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 0.48 & & & & \\ 0.10 & 0.47 & & & \\ 0.19 & 0.28 & 0.61 & & \\ 0.05 & 0.20 & 0.18 & 0.46 & \\ 0.10 & 0.08 & 0.10 & 0.14 & 0.43 \end{bmatrix}$$

$$F_{ML} = 0.098$$

การประมาณค่าจากความเป็นไปได้สูงสุด

- F_{ML} จะเป็นค่าที่บอกว่าเมทริกซ์ความแปรปรวนร่วมที่ได้จากโมเดล (Model-implied Covariance Matrix, **Σ**) นั้นแตกต่างจากเมทริกซ์ความแปรปรวนร่วมที่ได้จากข้อมูล (Sample Covariance Matrix, **S**) มากน้อยเพียงใด
- ยิ่งโมเดลสอดคล้องกับข้อมูลมากเพียงใด ค่า F_{ML} จะยิ่งน้อย
- สามารถทดสอบว่า **Σ** แตกต่างจาก **S** อย่างมีนัยสำคัญหรือไม่
  - ถ้าไม่ถึงระดับนัยสำคัญ แสดงว่ามีโอกาสที่ความแปรปรวนร่วมจากโมเดลสอดคล้องกับความแปรปรวนร่วมของข้อมูลในประชากร แสดงว่าโมเดลนี้อาจเป็นโมเดลที่อยู่ภายใต้ข้อมูล (โมเดลไม่เหมาะสมกับข้อมูล)
  - หากแตกต่างอย่างมีนัยสำคัญ แสดงว่าโมเดลนี้ไม่น่าจะเป็นโมเดลที่ทำให้เกิดข้อมูล (โมเดลเหมาะสมกับข้อมูล)

การประมาณค่าจากความเป็นไปได้สูงสุด

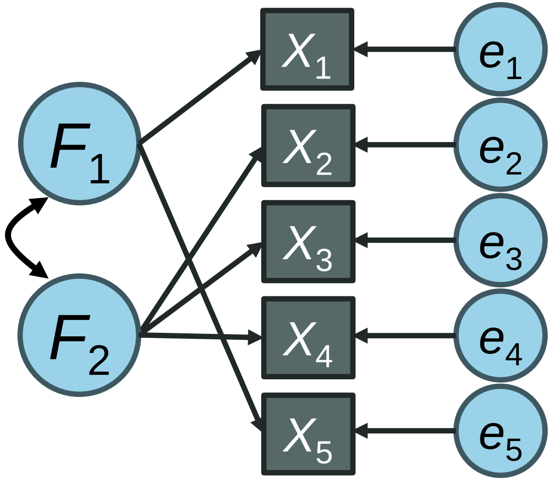
- ในการพิสูจน์ $H_0: \mathbf{\Sigma}_M = \mathbf{\Sigma}_S$ สามารถทำได้ด้วยสถิติการทดสอบ (Test Statistics) โดย $\mathbf{\Sigma}_S$ เป็นความแปรปรวนร่วมระหว่างตัวแปรในประชากรดังนี้

$$\chi^2 = (N-1)F_{ML}$$

- N = จำนวนกลุ่มตัวอย่าง
- ถ้าสมมติฐานว่าง (H_0) เป็นจริงแล้ว χ^2 จะมีการกระจายเป็นไคสแควร์
- ค่าองศาอิสระ (Degree of freedom, *df*) จะเท่ากับจำนวนข้อมูลในเมทริกซ์ความแปรปรวนร่วมของข้อมูลลบด้วยจำนวนพารามิเตอร์
  - จำนวนข้อมูลในเมทริกซ์ความแปรปรวนร่วม = $p(p+1)/2$ โดย p = จำนวนตัวบ่งชี้
  - สูตรในตำรามักเขียนว่า $df = \frac{p(p+1)}{2} - q$ โดย q = จำนวนพารามิเตอร์

การประมาณค่าจากความเป็นไปได้สูงสุด

- จากตัวอย่างมีจำนวนพารามิเตอร์ 11 ตัว
- จำนวนข้อมูล (5)(5 + 1)/2 = 15 ตัว
- *df* = (5)(5 + 1)/2 – 11 = 15 – 11 = 4



$$\mathbf{\Phi} = \begin{bmatrix} 1 & \\ 0.72 & 1 \end{bmatrix}$$

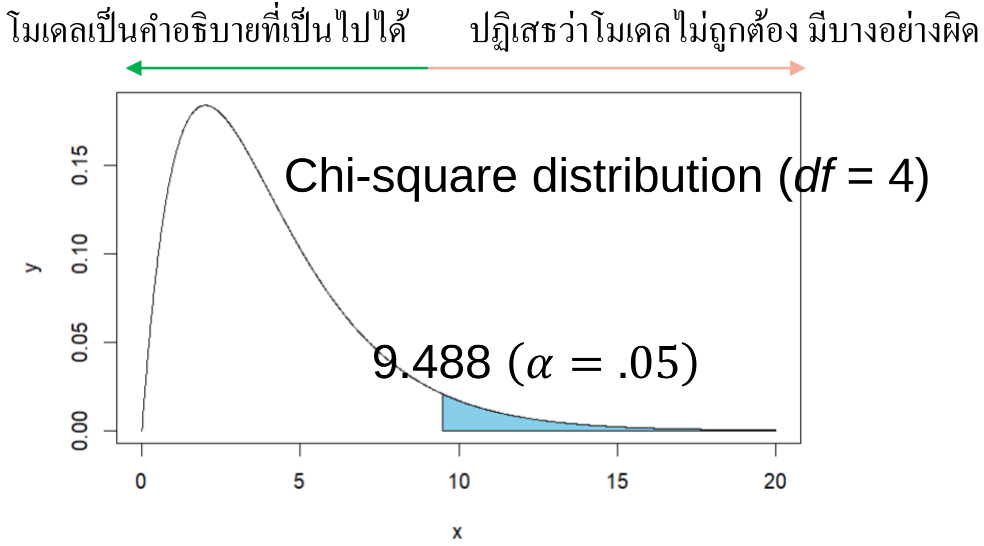
$$\mathbf{\Lambda} = \begin{bmatrix} 0.35 & 0 \\ 0 & 0.50 \\ 0 & 0.56 \\ 0 & 0.36 \\ 0.28 & 0 \end{bmatrix}$$

$$\mathbf{\Theta} = \begin{bmatrix} 0.36 & & & & \\ 0 & 0.22 & & & \\ 0 & 0 & 0.29 & & \\ 0 & 0 & 0 & 0.32 & \\ 0 & 0 & 0 & 0 & 0.35 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 0.48 & & & & \\ 0.10 & 0.47 & & & \\ 0.19 & 0.28 & 0.61 & & \\ 0.05 & 0.20 & 0.18 & 0.46 & \\ 0.10 & 0.08 & 0.10 & 0.14 & 0.43 \end{bmatrix}$$

การประมาณค่าจากความเป็นไปได้สูงสุด

- ค่าของ (N – 1)F_{ML} หรือ χ^2 จะมีการกระจายดังนี้ ถ้าสมมติฐานว่างเป็นจริง (โมเดลที่ตั้งไว้เป็นโมเดลสร้างข้อมูลในประชากร)



```
> m2 <- '
+ f1 =~ x1 + x5
+ f2 =~ x2 + x3 + x4
+ '
> out2 <- cfa(m2, data=dat, std.lv=TRUE)
> summary(out2)
lavaan 0.6-12 ended normally after 22 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        11

  Number of observations                           200

Model Test User Model:

  Test statistic                                19.550
  Degrees of freedom                                 4
  P-value (Chi-square)                           0.001
```

โมเดลสององค์ประกอบที่ได้ตั้งเป็นสมมติฐาน

ใช้คำสั่ง cfa ใน lavaan package เพื่อวิเคราะห์องค์ประกอบแบบยืนยัน

ใช้วิธีวิเคราะห์แบบความเป็นไปได้สูงสุด (Maximum Likelihood)

ใช้อัลกอริทึมในการหาความเป็นไปได้สูงสุดผ่านฟังก์ชั่น nlminb

จำนวนพารามิเตอร์

จำนวนกลุ่มตัวอย่าง

ค่าไคสแควร์

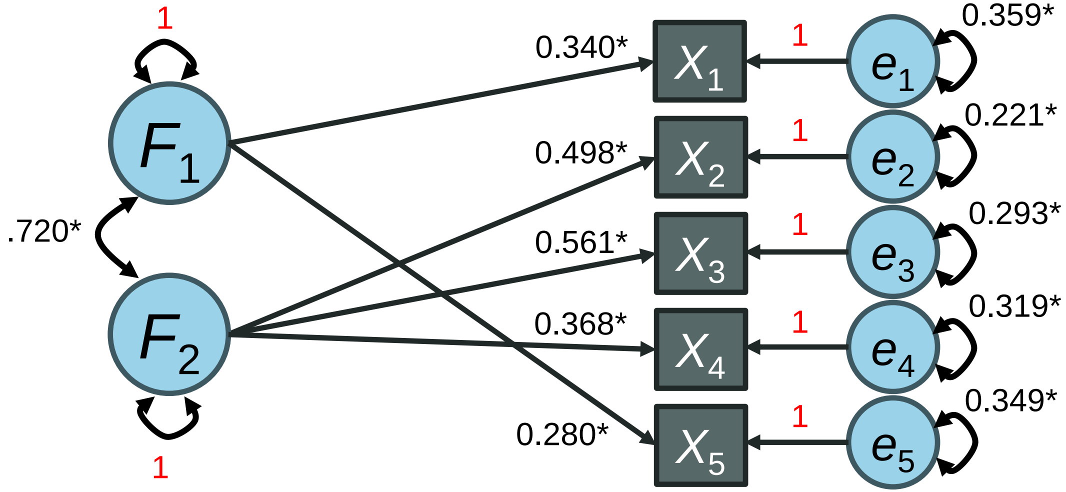
$df = 4$

$p = .001$ ปฏิเสธโมเดล

โดยปกติจะไม่ใช้ p ตัดสินโมเดล แต่ ณ เวลานี้ เป็นแบบนี้ไปก่อน

เครื่องหมาย =~ แสดงว่าองค์ประกอบอะไร ถูกอธิบายด้วยตัวบ่งชี้ (ตัวแปรสังเกตได้) อะไร
เครื่องหมาย ~~ แสดงถึงความแปรปรวนร่วม (Covariance) ระหว่างตัวแปรซ้ายและขวา
ถ้าตัวแปรซ้ายและขวาของเครื่องหมาย ~~ เป็นตัวแปรเดียวกัน จะหมายถึงความแปรปรวน (Variance)

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    x1                0.340    0.077    4.429    0.000
    x5                0.280    0.067    4.162    0.000
  f2 =~
    x2                0.498    0.053    9.354    0.000
    x3                0.561    0.061    9.267    0.000
    x4                0.368    0.052    7.094    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 ~~
    f2                0.720    0.140    5.158    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.359    0.055    6.496    0.000
   .x5                0.349    0.045    7.749    0.000
   .x2                0.221    0.040    5.546    0.000
   .x3                0.293    0.051    5.706    0.000
   .x4                0.319    0.038    8.430    0.000
    f1                1.000
    f2                1.000
```



ตัวเลขสีแดง คือ ค่าที่ถูกตรึงให้คงที่ในโมเดล สถิติไม่ได้ประมาณค่า
ตัวเลขสีดำ คือ ค่าที่ถูกประมาณค่าในโมเดล หรือค่าพารามิเตอร์
*p < .05 หมายถึง ค่าดังกล่าวแตกต่างจาก 0 อย่างมีนัยสำคัญ

การประมาณค่าจากความเป็นไปได้สูงสุด

- วิเคราะห์ข้อมูลมาตรวัดแรงจูงใจในการทำงาน (Manifest Needs Questionnaire)
- ข้อมูลจาก 799 คน
- ข้อคำถามแบ่งเป็น 4 องค์ประกอบ
  - Need for Achievement (Ach)
  - Need for Affiliation (Aff)
  - Need for Autonomy (Auto)
  - Need for Power (Pow)

คุณเป็นคนแบบไหนเวลาที่คุณทำงานร่วมกันคนอื่น?

		ไม่ตรงมากที่สุด	ไม่ตรง	ตรงและไม่ตรงพอ ๆ กัน	ตรง	ตรงมากที่สุด
ในการทำงาน ฉันชอบการทำเป้าหมายยาก ๆ (แต่เป็นไปได้) ให้สำเร็จ	Ach	○	○	○	○	○
การริเริ่มและรักษาความสัมพันธ์ที่เป็นมิตรกับเพื่อนร่วมงานทำให้ฉันพึงพอใจ	Aff	○	○	○	○	○
หากเลือกได้ ฉันชอบที่จะทำงานของฉันเอง โดยไม่ยุ่งเกี่ยวกับใคร	Auto	○	○	○	○	○
การได้มีบทบาทที่มีอำนาจเหนือเพื่อนร่วมงานเป็นเรื่องสำคัญสำหรับฉัน	Pow	○	○	○	○	○
ในการทำงาน ฉันชอบการทำงานตามเป้าหมายที่ชัดเจนและท้าทาย	Ach	○	○	○	○	○
ฉันชอบการรู้จักนิสัยและความสนใจของเพื่อนร่วมงาน	Aff	○	○	○	○	○
ในการทำงาน ฉันมองว่าการแก้ไขปัญหาด้วยตัวเอง โดยไม่ขอความช่วยเหลือจากคนอื่นเป็นเรื่องสำคัญ	Auto	○	○	○	○	○
หากเลือกได้ ฉันชอบเป็น "ผู้ออกคำสั่ง" ในการทำงาน	Pow	○	○	○	○	○
ฉันรู้สึกว่าการยืนหยัดในช่วงเวลายากลำบากเป็นสิ่งสำคัญในการทำงาน	Ach	○	○	○	○	○
ฉันใส่ใจความรู้สึกของเพื่อนร่วมงานด้วยความจริงใจ	Aff	○	○	○	○	○
หากเลือกได้ ฉันจะเลือกทำงานคนเดียวมากกว่าทำงานเป็นทีม	Auto	○	○	○	○	○
ฉันชอบการได้รับการยอมรับเรื่องการบริหารจัดการการกระทำของเพื่อนร่วมงาน	Pow	○	○	○	○	○
ฉันพึงพอใจเมื่ออยู่กับเพื่อนร่วมงาน	Aff	○	○	○	○	○
ฉันอยากรับรู้ว่าความสามารถของฉันทำให้ฉันประสบความสำเร็จ โดยไม่มีความช่วยเหลือจากเพื่อนร่วมงาน	Auto	○	○	○	○	○
ฉันรู้สึกมีคุณค่าเมื่อเพื่อนร่วมงานพึ่งพาฉัน	Pow	○	○	○	○	○

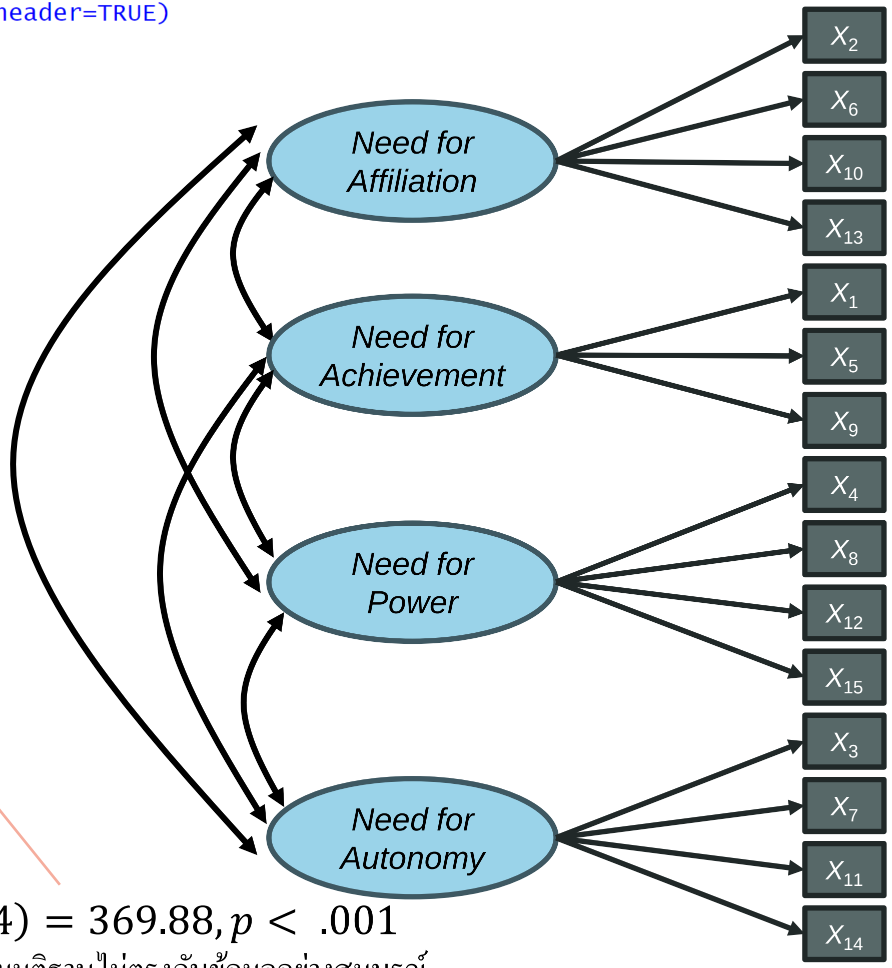
```
> datneed <- read.table("lecture8manifestneed.csv", sep=",", header=TRUE)
> mneed <- '
+ naff =~ x2 + x6 + x10 + x13
+ nach =~ x1 + x5 + x9
+ ndom =~ x4 + x8 + x12 + x15
+ nauto =~ x3 + x7 + x11 + x14
+ '
> outneed <- cfa(mneed, data=datneed, std.lv=TRUE)
> summary(outneed)
lavaan 0.6-12 ended normally after 25 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        36

  Number of observations                           799

Model Test User Model:

  Test statistic                               369.880
  Degrees of freedom                                84
  P-value (Chi-square)                           0.000
```

จำนวนพารามิเตอร์ 36 ตัว มาจาก

- ความแปรปรวนร่วมระหว่างองค์ประกอบ 6 ตัว
- น้ำหนักองค์ประกอบ 15 ตัว
- ความแปรปรวนของค่าความผิดพลาด 15 ตัว

จำนวนข้อมูล = 15(15 + 1)/2 = 120 ตัว

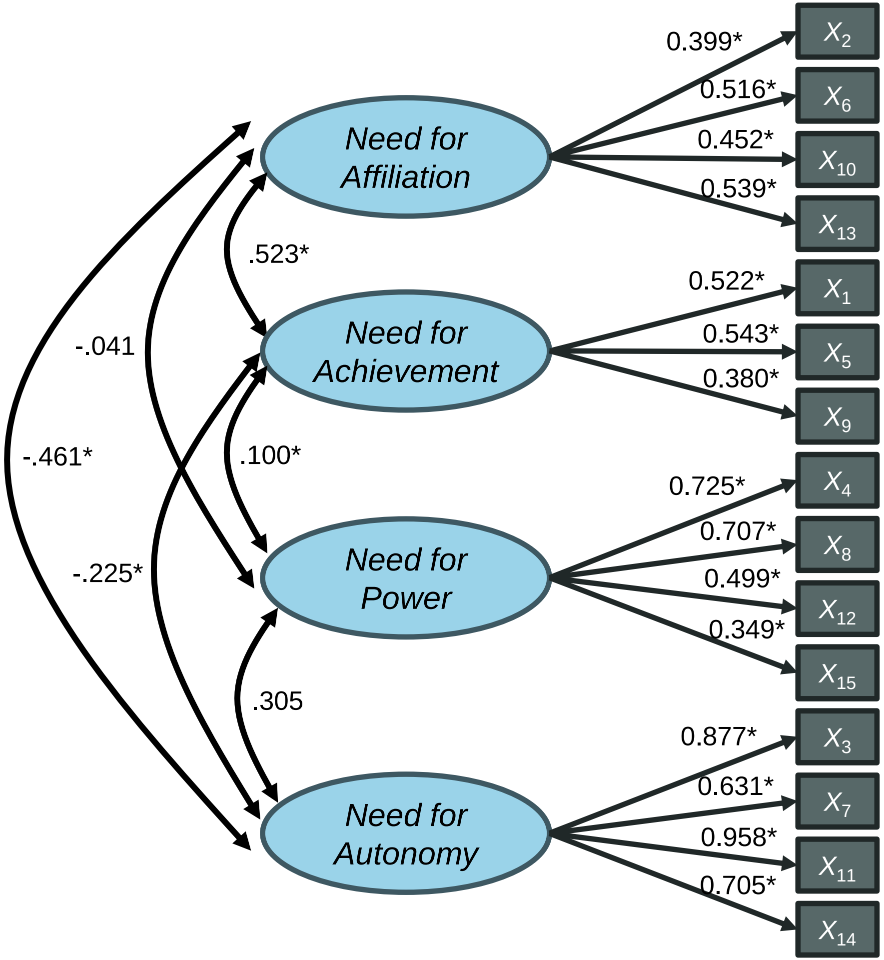
df = 120 – 36 = 84



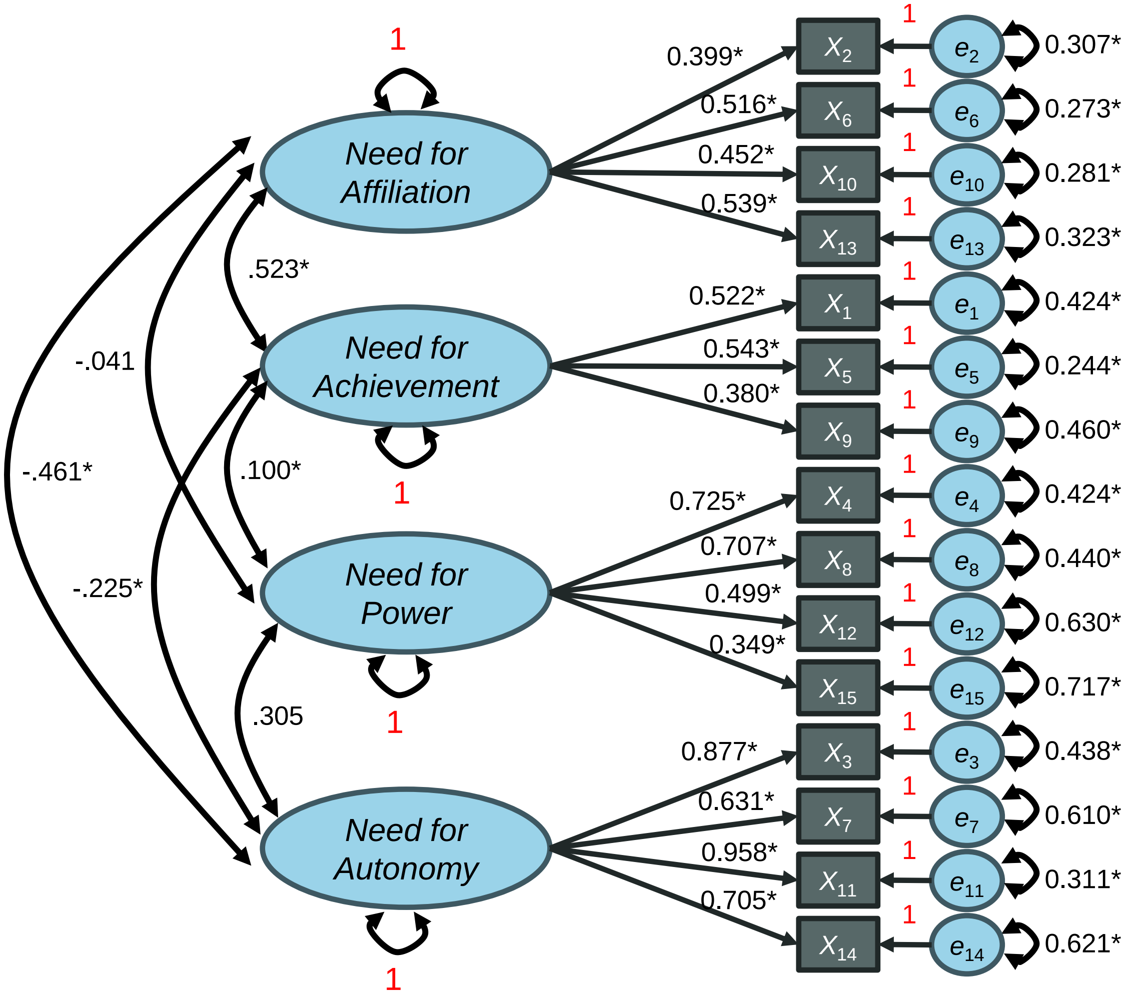
$$\chi^2(84) = 369.88, p < .001$$

โมเดลตามสมมติฐานไม่ตรงกับข้อมูลอย่างสมบูรณ์

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  naff =~
    x2                0.399    0.025   15.995    0.000
    x6                0.516    0.026   19.898    0.000
    x10               0.452    0.025   18.090    0.000
    x13               0.539    0.028   19.408    0.000
  nach =~
    x1                0.522    0.034   15.398    0.000
    x5                0.543    0.031   17.703    0.000
    x9                0.380    0.031   12.133    0.000
  ndom =~
    x4                0.725    0.037   19.529    0.000
    x8                0.707    0.037   19.166    0.000
    x12               0.499    0.036   13.938    0.000
    x15               0.349    0.036    9.669    0.000
  nauto =~
    x3                0.877    0.035   25.308    0.000
    x7                0.631    0.034   18.538    0.000
    x11               0.958    0.034   28.247    0.000
    x14               0.705    0.035   19.986    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  naff ~~
    nach              0.523    0.041   12.825    0.000
    ndom             -0.041    0.047   -0.867    0.386
    nauto            -0.461    0.037  -12.602    0.000
  nach ~~
    ndom              0.100    0.049    2.034    0.042
    nauto            -0.225    0.044   -5.061    0.000
  ndom ~~
    nauto             0.305    0.041    7.436    0.000
```

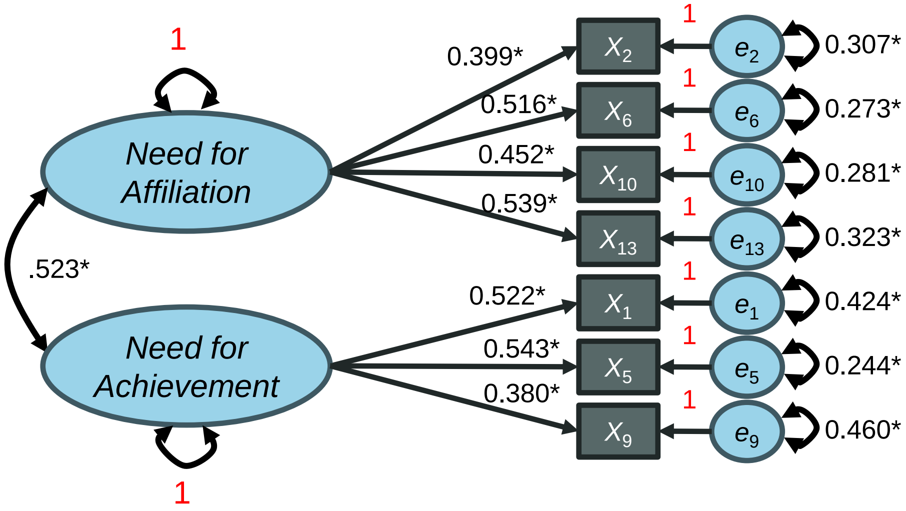


```
Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x2                0.307    0.018   17.022    0.000
   .x6                0.273    0.019   14.344    0.000
   .x10               0.281    0.018   15.803    0.000
   .x13               0.323    0.022   14.784    0.000
   .x1                0.424    0.031   13.827    0.000
   .x5                0.244    0.026    9.363    0.000
   .x9                0.460    0.027   17.249    0.000
   .x4                0.424    0.039   10.806    0.000
   .x8                0.440    0.038   11.440    0.000
   .x12               0.630    0.036   17.286    0.000
   .x15               0.717    0.038   18.857    0.000
   .x3                0.438    0.032   13.815    0.000
   .x7                0.610    0.034   17.886    0.000
   .x11               0.311    0.030   10.185    0.000
   .x14               0.621    0.036   17.382    0.000
    naff              1.000
    nach              1.000
    ndom              1.000
    nauto             1.000
```



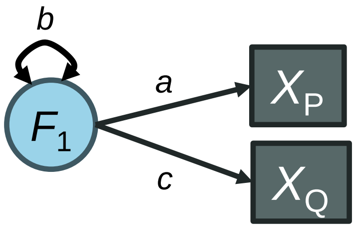
โมเดลองค์ประกอบแบบยืนยันนี้ แสดงถึงความสัมพันธ์ระหว่างตัวแปร

- ตัวแปรที่อยู่ในองค์ประกอบเดียวกันสัมพันธ์กันเนื่องจากอิทธิพลขององค์ประกอบที่มีต่อแต่ละตัวแปร



$$
\begin{aligned}
&Cov(X_2, X_6) \\
&= Cov(0.399F_1 + e_2, 0.516F_1 + e_6) \\
&= Cov(0.399F_1, 0.516F_1) + Cov(0.399F_1, e_6) + Cov(e_2, 0.516F_1) + Cov(e_2, e_6) \\
&= Cov(0.399F_1, 0.516F_1) \\
&= 0.399 \times 0.516 \times Cov(F_1, F_1) = 0.206 \times Var(F_1) = 0.206
\end{aligned}
$$

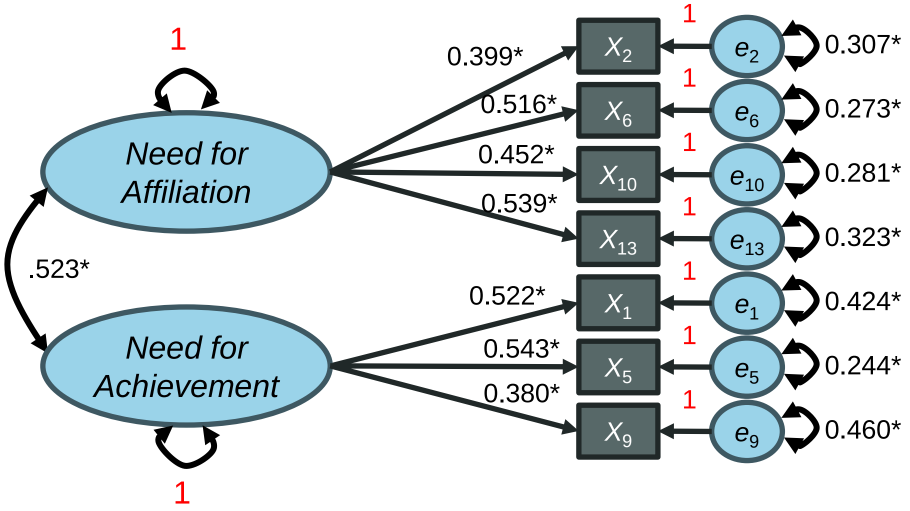
กฎการติดตาม (Tracing Rule)



$$Cov(X_P, X_Q) = a \times b \times c$$

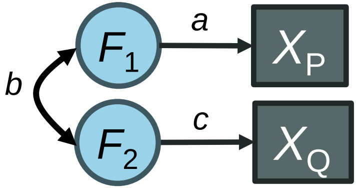
โมเดลองค์ประกอบแบบยืนยันนี้ แสดงถึงความสัมพันธ์ระหว่างตัวแปร

- ตัวแปรที่อยู่ต่างองค์ประกอบกันสัมพันธ์กันเนื่องจากความสัมพันธ์ระหว่างองค์ประกอบที่ต่างกัน และน้ำหนักองค์ประกอบที่มีต่อตัวแปร



$$
\begin{aligned}
&Cov(X_2, X_1)\\
&= Cov(0.399F_1 + e_2, 0.522F_2 + e_1)\\
&= Cov(0.399F_1, 0.522F_2) + Cov(0.399F_1, e_1) + Cov(e_2, 0.522F_2) + Cov(e_2, e_1)\\
&= Cov(0.399F_1, 0.522F_2)\\
&= 0.399 \times 0.522 \times Cov(F_1, F_2) = 0.208 \times 0.523 = 0.109
\end{aligned}
$$

กฎการติดตาม (Tracing Rule)



$$Cov(X_P, X_Q) = a \times b \times c$$

การประมาณค่าจากความเป็นไปได้สูงสุด

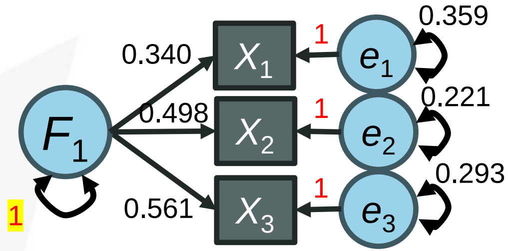
- กฎการติดตาม (Tracing Rule) เป็นวิธีการตรวจสอบว่าสมาชิกของ Model-implied Mean Vector and Covariance Matrix เกิดจากการประกอบกันของค่าพารามิเตอร์ใดในโมเดล
- เป็นวิธีการอย่างง่าย รวดเร็ว ไม่จำเป็นต้องมาคำนวณเมทริกซ์ดังสมการ

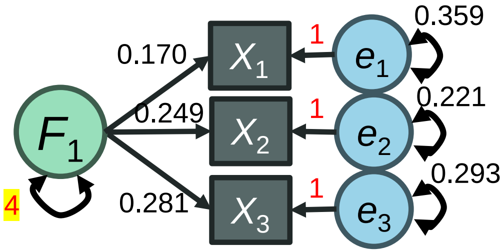
$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

- อย่างไรก็ตาม หากคำนวณค่าคาดหวังและความแปรปรวนร่วมของตัวแปรสุ่มเป็น ก็สามารถทดสอบกฎการติดตามได้ แม้จะไม่เร็วเท่า

การกำหนดขนาดองค์ประกอบ

- ในตัวอย่างที่ผ่านมา กำหนดให้ขนาดความแปรปรวนขององค์ประกอบเท่ากับ 1 ไม่ให้มีการประมาณค่า
- ในที่แท้จริง เราสามารถกำหนดขนาดความแปรปรวนเป็นเท่าไรก็ได้ โมเดลก็จะปรับขนาดน้ำหนักองค์ประกอบ แล้วยังได้ $\boldsymbol{\Sigma}$ ค่าเดิม





$$Var(X_1) = Var(0.340F_1 + e_1) = 0.340^2 Var(F_1) + Var(e_1) = 0.1156(1) + 0.359 = 0.477$$

$$Var(X_1) = Var(0.170F_1 + e_1) = 0.170^2 Var(F_1) + Var(e_1) = 0.0289(4) + 0.359 = 0.477$$

การกำหนดขนาดองค์ประกอบ

- ดังนั้นจะกำหนดขนาด (Scaling) องค์ประกอบ ให้มีขนาดใดก็ได้ แต่จำเป็นต้องกำหนดขนาดเริ่มต้น เพื่อให้โมเดลประมาณค่า หากไม่กำหนดขนาดเลย โมเดลจะไม่สามารถประมาณค่าได้ เพราะมีชุดคำตอบที่เป็นไปได้มากกว่า 1 ชุด
  - เหมือนการบอกทาง ที่จะบอกว่าให้เดินไป เลี้ยวซ้าย เลี้ยวขวา ต้องบอกว่าเริ่มจากที่ไหน
- วิธีมาตรฐานในการกำหนดขนาดมี 2 รูปแบบ คือ
  - การกำหนดความแปรปรวนองค์ประกอบ (Fixed Factor Approach) คือ กำหนดค่าความแปรปรวนขององค์ประกอบเท่ากับค่าใดค่าหนึ่งเลย โดยปกติจะกำหนดให้มีค่าเท่ากับ 1 เพื่อให้ความแปรปรวนร่วมเทียบเท่ากับค่าสหสัมพันธ์
  - การกำหนดขนาดน้ำหนักองค์ประกอบตัวบ่งชี้หลัก (Manifest Variable Approach หรือ Marker Variable Approach) คือ การกำหนดน้ำหนักองค์ประกอบของตัวบ่งชี้หนึ่งเท่ากับค่าคงที่ มักกำหนดให้มีค่าเท่ากับ 1 แล้วมาตรขององค์ประกอบจะเทียบเท่ากับมาตรของตัวบ่งชี้ดังกล่าว

กำหนดน้ำหนักองค์ประกอบตัวบ่งชี้หลัก

```
> m2mv1 <- '
+ f1 =~ 1*x1 + NA*x5
+ f2 =~ 1*x2 + NA*x3 + NA*x4
+ f1 ~~ NA*f1
+ f2 ~~ NA*f2
+ f1 ~~ NA*f2
+ '
> out2mv1 <- cfa(m2mv1, data=dat)
> summary(out2mv1)
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    x1                1.000
    x5                0.824    0.261    3.153    0.002
  f2 =~
    x2                1.000
    x3                1.126    0.171    6.576    0.000
    x4                0.740    0.124    5.969    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 ~~
    f2                0.122    0.031    3.899    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
    f1                0.116    0.052    2.214    0.027
    f2                0.248    0.053    4.677    0.000
   .x1                0.359    0.055    6.496    0.000
   .x5                0.349    0.045    7.749    0.000
   .x2                0.221    0.040    5.546    0.000
   .x3                0.293    0.051    5.706    0.000
   .x4                0.319    0.038    8.430    0.000
```

น้ำหนักองค์ประกอบเป็น 0.82 เท่าของตัวบ่งชี้หลัก

กำหนดให้คงที่

กำหนดให้คงที่

กำหนดความแปรปรวนองค์ประกอบ

```
> m2ff1 <- '
+ f1 =~ NA*x1 + NA*x5
+ f2 =~ NA*x2 + NA*x3 + NA*x4
+ f1 ~~ 1*f1
+ f2 ~~ 1*f2
+ f1 ~~ NA*f2
+ '
> out2ff1 <- cfa(m2ff1, data=dat)
> summary(out2ff1)
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    x1                0.340    0.077    4.429    0.000
    x5                0.280    0.067    4.162    0.000
  f2 =~
    x2                0.498    0.053    9.354    0.000
    x3                0.561    0.061    9.267    0.000
    x4                0.368    0.052    7.094    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 ~~
    f2                0.720    0.140    5.158    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
    f1                1.000
    f2                1.000
   .x1                0.359    0.055    6.496    0.000
   .x5                0.349    0.045    7.749    0.000
   .x2                0.221    0.040    5.546    0.000
   .x3                0.293    0.051    5.706    0.000
   .x4                0.319    0.038    8.430    0.000
```

NA* เป็นการกำหนดว่าเป็นพารามิเตอร์ที่จะประมาณค่า

1* บอกว่าค่านี้จะกำหนดไว้ว่าเท่ากับ 1 ไม่ประมาณค่า

องค์ประกอบเพิ่มขึ้น 1 ส่วนเบี่ยงเบนมาตรฐาน ตัวบ่งชี้ x1 มีค่าเพิ่มขึ้น 0.34 แต้ม

ค่าสหสัมพันธ์ระหว่างสององค์ประกอบ

กำหนดให้คงที่

ค่าความแปรปรวนของค่าคงเหลือของตัวบ่งชี้ ทั้งสองโมเดลเท่ากัน

กำหนดน้ำหนักองค์ประกอบตัวบ่งชี้หลัก

```
> m2mv1 <- '
+ f1 =~ 1*x1 + NA*x5
+ f2 =~ 1*x2 + NA*x3 + NA*x4
+ f1 ~~ NA*f1
+ f2 ~~ NA*f2
+ f1 ~~ NA*f2
+ '
> out2mv1 <- cfa(m2mv1, data=dat)
> summary(out2mv1)
```

```
> m2mv2 <- '
+ f1 =~ x1 + x5
+ f2 =~ x2 + x3 + x4
+ '
> out2mv2 <- cfa(m2mv2, data=dat)
```

ค่าเริ่มต้นของ lavaan จะกำหนดให้น้ำหนักองค์ประกอบของตัวบ่งชี้แรกคงที่มีค่าเท่ากับ 1 และตัวที่เหลือประมาณค่า และประมาณค่าความแปรปรวนขององค์ประกอบเช่นกัน

กำหนดความแปรปรวนองค์ประกอบ

```
> m2ff1 <- '
+ f1 =~ NA*x1 + NA*x5
+ f2 =~ NA*x2 + NA*x3 + NA*x4
+ f1 ~~ 1*f1
+ f2 ~~ 1*f2
+ f1 ~~ NA*f2
+ '
> out2ff1 <- cfa(m2ff1, data=dat)
> summary(out2ff1)
```

```
> m2ff2 <- '
+ f1 =~ x1 + x5
+ f2 =~ x2 + x3 + x4
+ '
> out2ff2 <- cfa(m2ff2, data=dat, std.lv=TRUE)
```

ถ้าใส่ `std.lv=TRUE` แล้ว จะเปลี่ยนการกำหนดความแปรปรวนองค์ประกอบให้เท่ากับ 1 แทน และประมาณค่าน้ำหนักองค์ประกอบทั้งหมด (ถ้าเป็นการวิเคราะห์โมเดลเชิงโครงสร้าง [SEM] ที่องค์ประกอบมีความสัมพันธ์เชิงสาเหตุ จะกำหนดความแปรปรวนคงเหลือเท่ากับ 1 แทน)

การกำหนดขนาดองค์ประกอบ

- ใน CFA วิธีที่ได้รับความนิยมคือการกำหนดความแปรปรวนองค์ประกอบ
- อย่างไรก็ตาม การกำหนดขนาดน้ำหนักองค์ประกอบตัวบ่งชี้หลัก จะเหมาะสมกับโมเดลที่ซับซ้อนมากขึ้น (เช่น การเปรียบเทียบคุณสมบัติมาตรวัดระหว่างกลุ่ม [Group Measurement Invariance]) เพราะการกำหนดโมเดลทำได้ง่ายกว่ามาก
  - ตัวบ่งชี้หลัก ควรเป็นตัวบ่งชี้ที่มีน้ำหนักองค์ประกอบสูง เพื่อให้โมเดลประมาณค่าได้ง่าย (ลู่เข้าสู่คำตอบ [Convergence] ได้ง่าย)
- สุดท้าย โมเดลเกือบทุกประเภท ไม่ว่าจะกำหนดขนาดองค์ประกอบรูปแบบใด นักวิจัยจะอ่านค่าพารามิเตอร์หลังจากที่ทำให้เป็นมาตรฐาน (Standardized parameters) แล้วอยู่ดี

ค่าพารามิเตอร์ที่ถูกทำให้เป็นมาตรฐาน

- ค่าพารามิเตอร์ที่ถูกทำให้เป็นมาตรฐาน (Standardized Parameters) จะช่วยทำให้การค่าพารามิเตอร์ทำได้ง่ายมากขึ้น
- หลักการคือ หากตัวแปรทั้งหมด ทั้งตัวบ่งชี้และองค์ประกอบ มีความแปรปรวนเท่ากับ 1 แล้ว ค่าพารามิเตอร์ในโมเดลจะมีค่าอย่างไร

จาก $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Theta}$

ให้ $\mathbf{D}_{\Sigma}^{-1/2} = \begin{bmatrix} 1/\sqrt{\sigma_1^2} & & & & \\ 0 & 1/\sqrt{\sigma_2^2} & & & \\ 0 & 0 & 1/\sqrt{\sigma_3^2} & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & 1/\sqrt{\sigma_p^2} \end{bmatrix}$ $\quad \mathbf{D}_{\Phi}^{-1/2} = \begin{bmatrix} 1/\sqrt{\phi_1} & & & & \\ 0 & 1/\sqrt{\phi_2} & & & \\ 0 & 0 & 1/\sqrt{\phi_3} & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & 1/\sqrt{\phi_k} \end{bmatrix}$

เพิ่มการหารส่วนเบี่ยงเบนมาตรฐาน

$$\mathbf{D}_{\Sigma}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}_{\Sigma}^{-1/2} = \mathbf{D}_{\Sigma}^{-1/2}\boldsymbol{\Lambda}\mathbf{D}_{\Phi}^{1/2}\mathbf{D}_{\Phi}^{-1/2}\boldsymbol{\Phi}\mathbf{D}_{\Phi}^{-1/2}\mathbf{D}_{\Phi}^{1/2}\boldsymbol{\Lambda}'\mathbf{D}_{\Sigma}^{-1/2} + \mathbf{D}_{\Sigma}^{-1/2}\boldsymbol{\Theta}\mathbf{D}_{\Sigma}^{-1/2}$$

ให้

$$\boldsymbol{\Sigma}^S = \mathbf{D}_{\Sigma}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}_{\Sigma}^{-1/2} \quad \boldsymbol{\Lambda}^S = \mathbf{D}_{\Sigma}^{-1/2}\boldsymbol{\Lambda}\mathbf{D}_{\Phi}^{1/2} \quad \boldsymbol{\Phi}^S = \mathbf{D}_{\Phi}^{-1/2}\boldsymbol{\Phi}\mathbf{D}_{\Phi}^{-1/2} \quad \boldsymbol{\Theta}^S = \mathbf{D}_{\Sigma}^{-1/2}\boldsymbol{\Theta}\mathbf{D}_{\Sigma}^{-1/2}$$

สรุป $\boldsymbol{\Sigma}^S = \boldsymbol{\Lambda}^S\boldsymbol{\Phi}^S\boldsymbol{\Lambda}^{S\prime} + \boldsymbol{\Theta}^S$

ค่าพารามิเตอร์ที่ถูกทำให้เป็นมาตรฐาน

- เมทริกซ์สหสัมพันธ์ระหว่างองค์ประกอบ $\mathbf{\Phi}^S$
  - สมาชิกบนแนวทแยงเป็น 1
  - สมาชิกนอกแนวทแยงจะเท่ากับสหสัมพันธ์ระหว่างองค์ประกอบ

$$Cor(F_m, F_n) = \frac{Cov(F_m, F_n)}{\sqrt{Var(F_m)Var(F_n)}}$$

ค่าพารามิเตอร์ที่ถูกทำให้เป็นมาตรฐาน

- เมทริกซ์น้ำหนักองค์ประกอบแบบเป็นมาตรฐาน (Standardized Loading Matrix) $\boldsymbol{\Lambda}^S$

$$\lambda_{ik}^S = \lambda_{ik}\sqrt{\frac{Var(F_k)}{Var(X_i)}}$$

  - น้ำหนักองค์ประกอบที่เป็นมาตรฐาน คือ หากองค์ประกอบหนึ่งเปลี่ยนแปลงไป 1 ส่วนเบี่ยงเบนมาตรฐานแล้วตัวบ่งชี้มีการเปลี่ยนแปลงไปกี่ส่วนเปลี่ยนแปลงมาตรฐาน
  - ส่วนใหญ่จะมีค่าระหว่าง -1 ถึง 1 แต่อาจน้อยกว่า -1 หรือมากกว่า 1 ได้ในกรณีพิเศษ

ค่าพารามิเตอร์ที่ถูกทำให้เป็นมาตรฐาน

- เมทริกซ์ความแปรปรวนร่วมของค่าคงเหลือแบบมีมาตรฐาน $\boldsymbol{\Theta}^S$
- สังเกตว่าผมไม่ใช้คำว่าสหสัมพันธ์ เนื่องจากตัวหาร เป็นส่วนเบี่ยงเบนมาตรฐานของตัวบ่งชี้ ไม่ใช่ส่วนเบี่ยงเบนมาตรฐานของค่าคงเหลือ

$$\theta_i^S = \theta_i / Var(X_i)$$

  - สมาชิกบนแนวทแยง เป็นสัดส่วนความแปรปรวนของค่าคงเหลือเทียบกับความแปรปรวนของตัวแปรทั้งหมด
  - ความแปรปรวนของค่าคงเหลือเทียบเท่ากับกี่เปอร์เซนต์ของความแปรปรวนของตัวบ่งชี้นั้น มีค่าที่เป็นไปได้ระหว่าง 0 (ไม่เหลือค่าคงเหลือ) ถึง 1 (องค์ประกอบอธิบายไม่ได้เลย)
  - เรียกค่านี้ว่า สัดส่วนจำเพาะ (Uniqueness)

ค่าพารามิเตอร์ที่ถูกทำให้เป็นมาตรฐาน

- เมทริกซ์ความแปรปรวนร่วมของค่าคงเหลือแบบมีมาตรฐาน $\boldsymbol{\Theta}^S$

$$\theta_{ij}^S = \theta_{ij}/\sqrt{Var(X_i)Var(X_j)}$$

  - สมาชิกนอกแนวทแยง ไม่ใช่สหสัมพันธ์ระหว่างค่าคงเหลือ
  - แต่เป็นความแปรปรวนร่วมของค่าคงเหลือสองตัวแปร ในหน่วยของตัวบ่งชี้เดิม
  - แปลความหมายได้ว่า หากตัวบ่งชี้หนึ่งเพิ่มขึ้น 1 ส่วนเบี่ยงเบนมาตรฐาน ค่าคงเหลือของอีกตัวบ่งชี้หนึ่งจะเปลี่ยนแปลงเทียบเท่ากับกี่ส่วนเบี่ยงเบนมาตรฐาน
  - ค่านี้จะเท่ากับ 1 ได้ เมื่อองค์ประกอบไม่สามารถอธิบายตัวบ่งชี้ได้เลย และค่าคงเหลือของทั้งสองตัวแปรเท่ากับ 1
  - ควรหลีกเลี่ยงการแปลความหมายค่านี้

```
> m2 <- '
+ f1 =~ x1 + x5
+ f2 =~ x2 + x3 + x4
+ '
> out2 <- cfa(m2, data=dat, std.lv=TRUE)
> summary(out2, std=TRUE)
```

ใส่ std=TRUE เพื่อแสดงสัมประสิทธิ์มาตรฐาน

เปลี่ยนให้องค์ประกอบมีความแปรปรวนเท่ากับ 1 เท่านั้น (สูตรใส่เฉพาะส่วนที่เกี่ยวข้องกับองค์ประกอบ) (ผมมักไม่อ่านคอลัมน์นี้)

เปลี่ยนให้ทั้งองค์ประกอบและตัวบ่งชี้มีความแปรปรวนเท่ากับ 1

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  f1 =~
    x1                0.340    0.077    4.429    0.000    0.340    0.494
    x5                0.280    0.067    4.162    0.000    0.280    0.428
  f2 =~
    x2                0.498    0.053    9.354    0.000    0.498    0.727
    x3                0.561    0.061    9.267    0.000    0.561    0.720
    x4                0.368    0.052    7.094    0.000    0.368    0.546
```

องค์ประกอบเปลี่ยนแปลง 1 *SD* ตัวบ่งชี้เปลี่ยนแปลง 0.34 หน่วย

องค์ประกอบเปลี่ยนแปลง 1 *SD* ตัวบ่งชี้เปลี่ยนแปลง 0.49 *SD*

หากน้ำหนักองค์ประกอบมาตรฐานต่ำกว่า .3 อาจมองได้ว่าตัวบ่งชี้นั้นได้รับอิทธิพลน้อยมากจากองค์ประกอบนั้น

ในองค์ประกอบที่ 2 พบว่า X_2 มีน้ำหนักสูงสุด รองลงมาคือ X_3 และ X_4

ค่าสหสัมพันธ์เท่ากับค่าความแปรปรวนร่วมที่ประมาณค่าจากโมเดล เนื่องจากกำหนดสเกลด้วยความแปรปรวนขององค์ประกอบเป็น 1

ค่าสหสัมพันธ์ระหว่างองค์ประกอบ

```
Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  f1 ~~
    f2                0.720    0.140    5.158    0.000    0.720    0.720

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .x1                0.359    0.055    6.496    0.000    0.359    0.756
   .x5                0.349    0.045    7.749    0.000    0.349    0.817
   .x2                0.221    0.040    5.546    0.000    0.221    0.471
   .x3                0.293    0.051    5.706    0.000    0.293    0.482
   .x4                0.319    0.038    8.430    0.000    0.319    0.701
    f1                1.000                               1.000    1.000
    f2                1.000                               1.000    1.000
```

สัดส่วนของความแปรปรวนคงเหลือเมื่อเทียบกับความแปรปรวนของตัวบ่งชี้เท่ากับ .756

แสดงว่ามีสัดส่วนที่องค์ประกอบสามารถอธิบายได้เท่ากับ .244

หากมีองค์ประกอบเดียวชี้หาตัวแปรหนึ่ง สามารถคำนวณสัดส่วนความคงเหลือผ่านน้ำหนักองค์ประกอบมาตรฐานได้เลย เช่น $.756 = 1 - (.340)^2$

```
> standardizedSolution(out2)
   lhs op rhs est.std    se      z pvalue ci.lower ci.upper
1   f1 =~  x1   0.494 0.104  4.729      0    0.289    0.698
2   f1 =~  x5   0.428 0.097  4.421      0    0.238    0.618
3   f2 =~  x2   0.727 0.059 12.335      0    0.612    0.843
4   f2 =~  x3   0.720 0.059 12.187      0    0.604    0.835
5   f2 =~  x4   0.546 0.064  8.493      0    0.420    0.672
6   x1 ~~  x1   0.756 0.103  7.338      0    0.554    0.958
7   x5 ~~  x5   0.817 0.083  9.844      0    0.654    0.979
8   x2 ~~  x2   0.471 0.086  5.498      0    0.303    0.639
9   x3 ~~  x3   0.482 0.085  5.673      0    0.316    0.649
10  x4 ~~  x4   0.701 0.070  9.978      0    0.564    0.839
11  f1 ~~  f1   1.000 0.000     NA     NA    1.000    1.000
12  f2 ~~  f2   1.000 0.000     NA     NA    1.000    1.000
13  f1 ~~  f2   0.720 0.140  5.158      0    0.447    0.994
```

หากต้องการทดสอบว่าค่ามาตรฐานของค่าพารามิเตอร์แตกต่างจาก 0 (หรือค่าอื่น) หรือไม่ ให้ใช้ค่าจาก standardizedSolution

การทดสอบจะไม่ได้ทดสอบค่าพารามิเตอร์อย่างเดียว จะคำนึงถึงความไม่แน่นอนของค่าความแปรปรวนขององค์ประกอบหรือตัวบ่งชี้ด้วย ทำให้การทดสอบตรงนี้จะถูกต้องมากกว่าการใช้การทดสอบค่าพารามิเตอร์ธรรมดา (แม้ว่าค่าส่วนใหญ่จะออกมาใกล้เคียงกัน)

สัดส่วนร่วม

- ในการวิเคราะห์ถดถอย สัดส่วนของความแปรปรวนของตัวแปรตามที่ตัวแปรต้นสามารถอธิบายได้ จะเรียกว่า สัมประสิทธิ์การทำนาย (Coefficient of Determination; R^2)
- ในการวิเคราะห์องค์ประกอบ สัดส่วนความแปรปรวนของตัวบ่งชี้แต่ละตัว ที่องค์ประกอบสามารถอธิบายได้ จะเรียกว่า สัดส่วนร่วม (Communalities; h^2) และส่วนที่องค์ประกอบไม่สามารถอธิบายได้จะเรียกว่า สัดส่วนจำเพาะ (Specificity; u^2)
- จาก

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

$$h_i^2 = \frac{(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}')_{ii}}{(\boldsymbol{\Sigma})_{ii}} = 1 - \frac{(\boldsymbol{\Theta})_{ii}}{(\boldsymbol{\Sigma})_{ii}} = 1 - \theta_i^S$$

สัดส่วนร่วม

- หากสัดส่วนร่วมมีค่าสูง แสดงว่าองค์ประกอบในโมเดลสามารถอธิบายตัวบ่งชี้แต่ละตัวได้ดี
- สามารถกำหนด `rsquare=TRUE` ในคำสั่ง `summary` เพื่อหาสัดส่วนร่วม

```
> m2 <- '
+ f1 =~ x1 + x5
+ f2 =~ x2 + x3 + x4
+ '
> out2 <- cfa(m2, data=dat, std.lv=TRUE)
> summary(out2, rsquare=TRUE)
```

```
R-Square:
                   Estimate
    x1                0.244
    x5                0.183
    x2                0.529
    x3                0.518
    x4                0.299
```

องค์ประกอบแรกอธิบายตัวบ่งชี้ที่ 1 และ 5 ได้น้อย เมื่อเทียบกับองค์ประกอบที่สองอธิบายตัวบ่งชี้ที่ 2, 3, และ 4

หากมีองค์ประกอบที่มาอธิบายตัวบ่งชี้หนึ่งมากกว่าหนึ่งองค์ประกอบ (จะกล่าวถึงในเรื่องน้ำหนักองค์ประกอบข้าม) สัดส่วนร่วมจะบอกการอธิบายจากองค์ประกอบทั้งหมด

การระบุโมเดล

- ในโมเดลที่ผ่านมา ตัวบ่งชี้แต่ละตัวจะมีองค์ประกอบที่มีอิทธิพลเพียงแค่องค์ประกอบเดียว และค่าคงเหลือไม่มีสหสัมพันธ์กันเลย จะเรียกโมเดลนี้ว่าโครงสร้างอย่างง่ายมาก (Very Simple Structure)
- แต่ในสภาพความเป็นจริง ตัวบ่งชี้บางตัวอาจได้รับอิทธิพลจากมากกว่าหนึ่งองค์ประกอบ เช่น

ความชอบพอ (Liking)

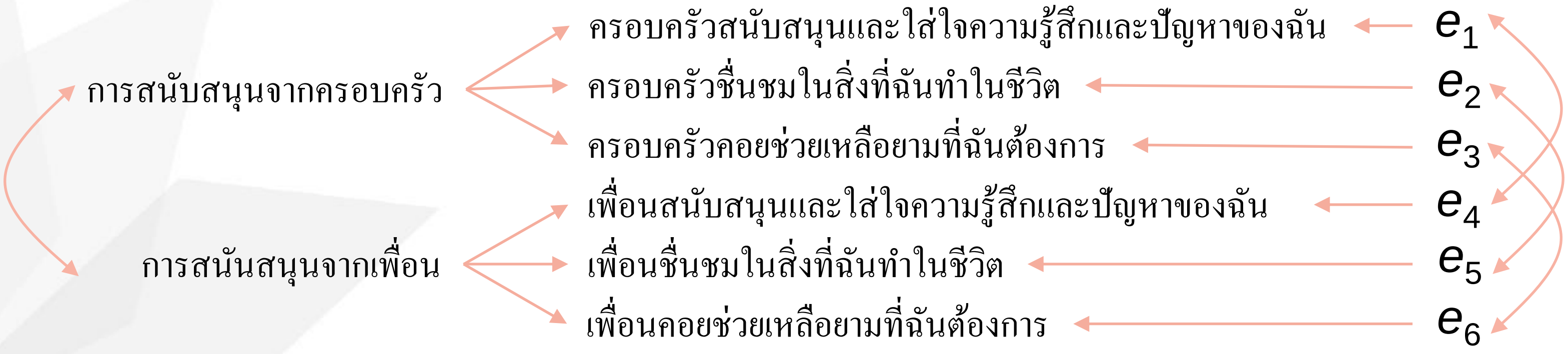
ความหลงใหล (Infatuation)

ฉันโปรดเธอ
ฉันชอบเธอ
ฉันรักเธอ
ฉันคลั่งเธอ
ฉันปลื้มเธอ
ฉันหลงใหลเธอ

คำว่าฉันรักเธอได้รับอิทธิพลจากทั้งความชอบและความหลงใหล ถึงจะเรียกว่ารักได้

การระบุโมเดล

- การที่ตัวบ่งชี้มีองค์ประกอบที่ส่งผลมากกว่า 1 องค์ประกอบ จะเรียกว่าน้ำหนักองค์ประกอบข้าม (Cross loadings) ซึ่งนักวิเคราะห์สามารถตั้งสมมติฐานในโมเดลให้มีน้ำหนักองค์ประกอบข้ามได้ ถ้า “ทฤษฎี” หรือมีคำอธิบายเป็นเหตุผลรองรับ
- นอกจากนี้ ค่าคงเหลือของตัวบ่งชี้ ซึ่งเป็นส่วนที่ไม่ได้อธิบายด้วยองค์ประกอบ อาจมีความสัมพันธ์ระหว่างกัน เช่น



การระบุโมเดล

- ความแปรปรวนร่วมระหว่างค่าคงเหลือ (Residual Covariance) จะเกิดขึ้นเมื่อองค์ประกอบและความสัมพันธ์ระหว่างองค์ประกอบ ไม่สามารถอธิบายความสัมพันธ์ระหว่างตัวบ่งชี้ได้ทั้งหมด
- นักวิเคราะห์สามารถใส่ความแปรปรวนร่วมระหว่างค่าคงเหลือได้ หากมีทฤษฎีหรือคำอธิบายความแปรปรวนร่วมระหว่างค่าคงเหลือดังกล่าว
- อย่างไรก็ตาม นักวิจัยไม่สามารถใส่น้ำหนักองค์ประกอบข้าม หรือความแปรปรวนค่าคงเหลือที่เป็นไปได้ทั้งหมดได้ บางโมเดลจะไม่สามารถประมาณค่าได้เพราะจำนวนพารามิเตอร์มากเกินไป
- หลักการระบุโมเดล (Model Identification) จะเป็นแนวคิดคร่าว ๆ ว่าโมเดลแบบใดสามารถประมาณค่าได้

การระบุโมเดล

- การระบุภาพรวมของโมเดล จะมีกฎ 2 กฎ คือ
  - กฎจำเป็น (Necessary rule หรือ Necessary-but-insufficient identification) คือ องศาอิสระ (*df*) ต้องมากกว่าหรือเท่ากับ 0 และองค์ประกอบทุกตัวต้องมีการกำหนดขนาด
  - กฎเพียงพอ (Sufficient rule) ซึ่งใช้ Rank ของเมทริกซ์เพื่อพิจารณาว่าการระบุโมเดลเหมาะสมหรือไม่ ซึ่งค่อนข้างซับซ้อน
- สรุปแบบง่าย คือ หากโมเดลไม่มีน้ำหนักองค์ประกอบข้าม (ตัวบ่งชี้แต่ละตัวมีองค์ประกอบเดียว) และความสัมพันธ์ระหว่างค่าคงเหลือ ซึ่งเรียกว่าโครงสร้างแบบง่าย (Simple Structure) โมเดลจะประมาณค่าได้เมื่อ
  - มีตัวบ่งชี้ 3 ตัวขึ้นไป หากมีองค์ประกอบเดียว
  - มีตัวบ่งชี้ 2 ตัวขึ้นไป หากมีตั้งแต่ 2 องค์ประกอบขึ้นไป

ตรงกับกฎที่ $df \geq 0$

ตัวบ่งชี้ต่อองค์ประกอบ = 2

จำนวนองค์ประกอบ	จำนวนข้อมูล	จำนวนพารามิเตอร์	*df*
1	3	4	-1
2	10	9	1
3	21	15	6
4	36	22	14
5	55	30	25

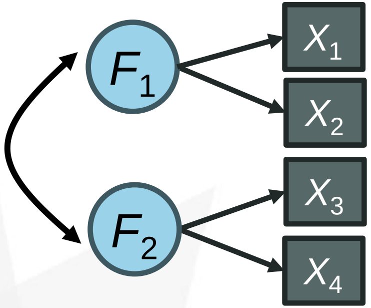
ตัวบ่งชี้ต่อองค์ประกอบ = 3

จำนวนองค์ประกอบ	จำนวนข้อมูล	จำนวนพารามิเตอร์	*df*
1	6	6	0
2	21	13	8
3	45	21	24
4	78	30	48
5	120	40	80

- ไม่นับโครงสร้างค่าเฉลี่ย (Mean structure)
- จำนวนข้อมูล $= p(p+1)/2$
- จำนวนพารามิเตอร์ $= 2p + K(K-1)/2$
- $df =$ จำนวนข้อมูล - จำนวนพารามิเตอร์

การระบุโมเดล

- แม้ว่าการระบุภาพรวมของโมเดลจะถูกต้อง แต่บางครั้งข้อมูลในโมเดลไม่เพียงพอในการประมาณค่าพารามิเตอร์ เรียกว่า การระบุโมเดลประจำจุด (Local Identification) เช่น



ตามหลักเกณฑ์แล้ว โมเดลนี้สามารถระบุได้ (Identified) เนื่องจากมี 2 องค์ประกอบ มีตัวบ่งชี้องค์ประกอบละ 2 ตัว ทำให้ *df* > 0 แต่หากดูการหาเมทริกซ์ความแปรปรวนจากโมเดล

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \end{bmatrix} \qquad \mathbf{\Phi} = \begin{bmatrix} 1 & \phi_{21} \\ \phi_{21} & 1 \end{bmatrix} \qquad \mathbf{\Theta} = \begin{bmatrix} \theta_{11} & & & \\ 0 & \theta_{22} & & \\ 0 & 0 & \theta_{33} & \\ 0 & 0 & 0 & \theta_{44} \end{bmatrix}$$

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta} = \begin{bmatrix} \lambda_{11}^2 + \theta_{11} & & & \\ \lambda_{21}\lambda_{11} & \lambda_{22}^2 + \theta_{22} & & \\ \lambda_{32}\phi_{21}\lambda_{11} & \lambda_{32}\phi_{21}\lambda_{21} & \lambda_{22}^2 + \theta_{33} & \\ \lambda_{42}\phi_{21}\lambda_{11} & \lambda_{32}\phi_{21}\lambda_{21} & \lambda_{42}\lambda_{32} & \lambda_{44}^2 + \theta_{44} \end{bmatrix}$$

การระบุโมเดล

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T + \boldsymbol{\Theta} = \begin{bmatrix} \lambda_{11}^2 + \theta_{11} & & & \\ \lambda_{21}\lambda_{11} & \lambda_{22}^2 + \theta_{22} & & \\ \lambda_{32}\boldsymbol{\phi}_{21}\lambda_{11} & \lambda_{32}\boldsymbol{\phi}_{21}\lambda_{21} & \lambda_{22}^2 + \theta_{33} & \\ \lambda_{42}\boldsymbol{\phi}_{21}\lambda_{11} & \lambda_{32}\boldsymbol{\phi}_{21}\lambda_{21} & \lambda_{42}\lambda_{32} & \lambda_{44}^2 + \theta_{44} \end{bmatrix}$$

ถ้าตัวสีน้ำเงินมีค่าใกล้ 0 (อาจเกิดจาก $\boldsymbol{\phi}_{21}$ มีค่าใกล้ 0) จะทำให้

1. การประมาณค่าสีแดง มีข้อมูล 3 ตัว คือ s_{11}, s_{22}, s_{21} แต่มีพารามิเตอร์ 4 ตัวคือ $\lambda_{11}, \lambda_{21}, \theta_{11}, \theta_{22}$
2. การประมาณค่าสีเขียว มีข้อมูล 3 ตัว คือ s_{33}, s_{44}, s_{32} แต่มีพารามิเตอร์ 4 ตัวคือ $\lambda_{32}, \lambda_{42}, \theta_{33}, \theta_{44}$

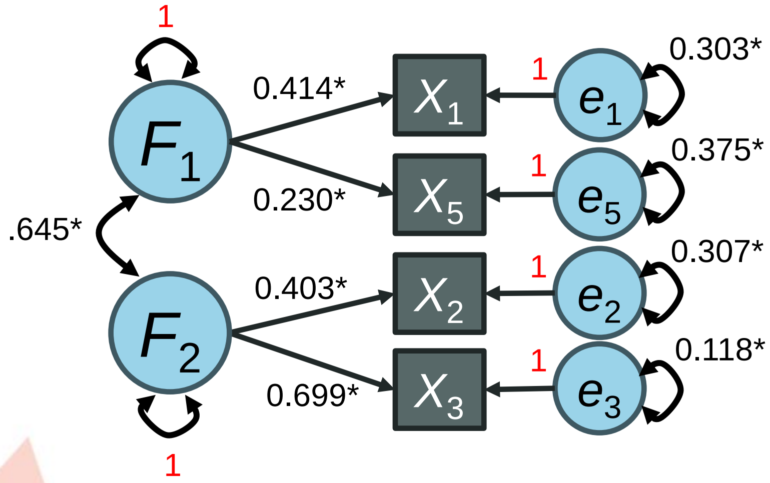
จากภาพนี้ แสดงให้เห็นว่า แม้โมเดลภาพรวมจะสามารถระบุได้ (Globally Identified) แต่ในรายละเอียดย่อยอาจจะไม่สามารถระบุได้ (Locally Unidentified) ซึ่งเป็นข้อควรระวังในการสร้างองค์ประกอบที่มี 2 ตัวบ่งชี้ เพราะต้องใช้ข้อมูลจากตัวแปรอื่น กล่าวคือ องค์ประกอบที่มี 2 ตัวบ่งชี้ต้องมีความสัมพันธ์ระดับหนึ่ง (เช่น .5 ขึ้นไป) กับองค์ประกอบอื่นในโมเดลเพื่อให้โมเดลประมาณค่าพารามิเตอร์ได้

การระบุโมเดล

- สรุป ถ้าโมเดลมีหลายองค์ประกอบ ใช้ตัวบ่งชี้ 2 ตัวขึ้นไปต่อองค์ประกอบ
- องค์ประกอบใดที่มี 2 ตัวบ่งชี้ ต้องมีความสัมพันธ์กับองค์ประกอบ (หรือตัวแปร) อื่น เพื่อช่วยในการประมาณค่า

องค์ประกอบที่มีตัวบ่งชี้เดียว

- องค์ประกอบไม่สามารถมีตัวชี้วัดเดียวได้ ไม่ว่ากรณีใดก็ตาม
  - เชิงแนวคิด ว่าเราไม่สามารถแยกระหว่างคะแนนองค์ประกอบและค่าผิดพลาดในการวัดออกจากกันได้
  - เชิงสถิติ องค์ประกอบที่มีตัวชี้วัดเดียวจะไม่สามารถระบุโมเดลได้ เพราะมีข้อมูลความแปรปรวนขององค์ประกอบอย่างเดียว ในการประมาณค่าทั้งน้ำหนักองค์ประกอบและความแปรปรวนค่าคงเหลือ
- อย่างไรก็ตาม บางครั้งนักวิเคราะห์ต้องการใช้ตัวบ่งชี้ไปหาความสัมพันธ์กับองค์ประกอบในโมเดล เช่น การสนับสนุนทางสังคมด้านต่างๆ สัมพันธ์กับความพึงพอใจในหรือไม่ โดยความพึงพอใจในชีวิตวัดด้วยข้อคำถามเดียว

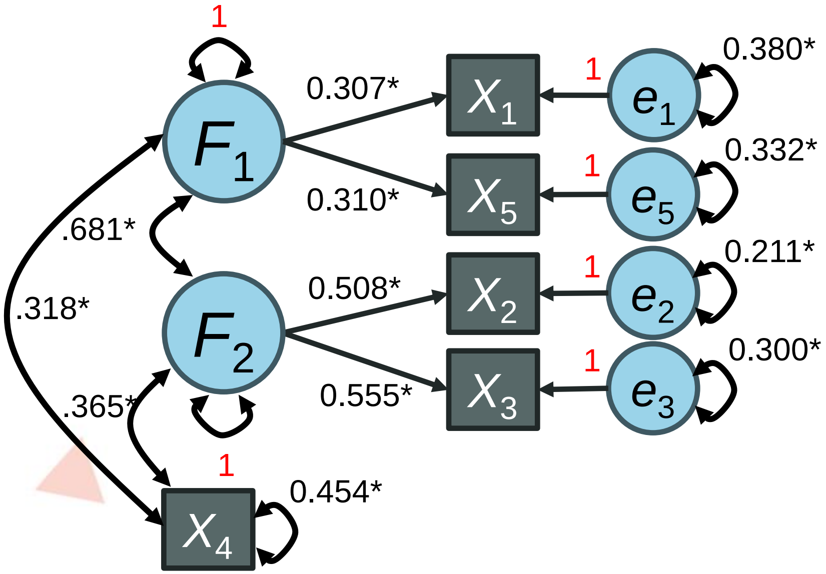


โมเดลตั้งต้น คือ โมเดล 2 องค์ประกอบแต่ละองค์ประกอบมี 2 ตัวบ่งชี้

```
> msingle0 <- '
+ f1 =~ x1 + x5
+ f2 =~ x2 + x3
+ '
> outsingle0 <- cfa(msingle0, data=dat, std.lv=TRUE)
> summary(outsingle0)
lavaan 0.6-12 ended normally after 24 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         9

  Number of observations                           200

Model Test User Model:

  Test statistic                                 0.851
  Degrees of freedom                                 1
  P-value (Chi-square)                           0.356

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    x1                0.414    0.091    4.553    0.000
    x5                0.230    0.063    3.653    0.000
  f2 =~
    x2                0.403    0.064    6.279    0.000
    x3                0.699    0.091    7.646    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 ~~
    f2                0.645    0.145    4.438    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.303    0.073    4.167    0.000
   .x5                0.375    0.043    8.795    0.000
   .x2                0.307    0.048    6.334    0.000
   .x3                0.118    0.114    1.041    0.298
    f1                1.000
    f2                1.000
```

โมเดล 2 องค์ประกอบ องค์ประกอบละ 2 ตัวบ่งชี้ สามารถประมาณค่าได้ หากค่าสหสัมพันธ์ระหว่างองค์ประกอบสูง

ค่าน้ำหนักองค์ประกอบ สหสัมพันธ์ระหว่างองค์ประกอบ และขนาดความแปรปรวนของค่าคงเหลือ ไม่เท่าเดิม

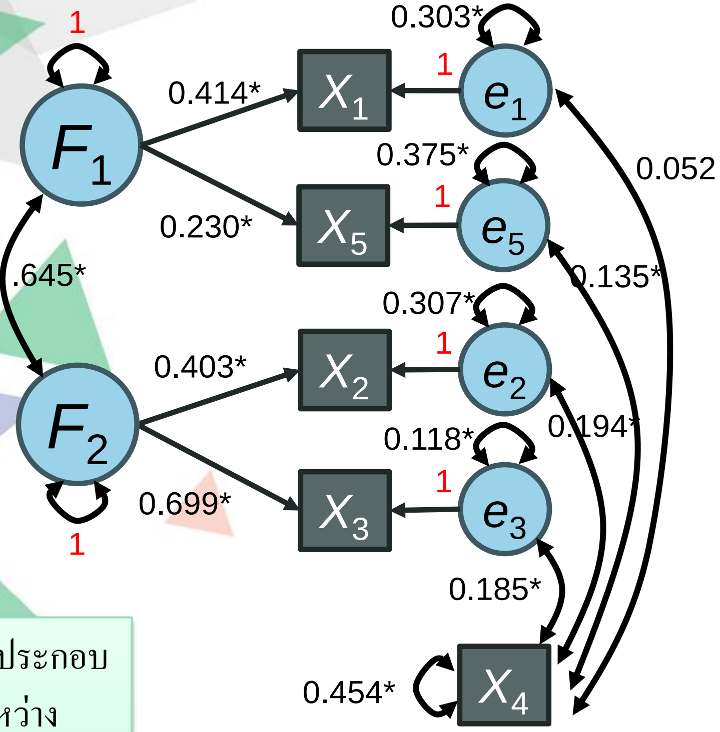


โมเดลนี้ เอา X_4 มาหาความสัมพันธ์กับ F_1 และ F_2

สังเกตว่า df สูงขึ้น เนื่องจากมีจำนวนข้อมูลที่เพิ่มขึ้น 5 ตัว คือ $s_{44}, s_{14}, s_{24}, s_{34}, s_{54}$ แต่มีค่าพารามิเตอร์ที่ประมาณค่าเพิ่มขึ้น 3 ตัว คือ $\phi_{31}, \phi_{32}, \phi_{33}$

มองอีกนัยหนึ่ง คือ ความสัมพันธ์ระหว่าง X_4 กับ X_1 และ X_4 กับ X_5 ถูกโมเดลให้เป็นความสัมพันธ์ผ่าน F_1 ทั้งคู่

```
> outsingle1 <- cfa(msingle1, data=dat, std.lv=TRUE)
> summary(outsingle1)
lavaan 0.6-12 ended normally after 27 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        12

  Number of observations                           200

Model Test User Model:

  Test statistic                                19.066
  Degrees of freedom                                 3
  P-value (Chi-square)                           0.000

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    x1                0.307    0.072    4.252    0.000
    x5                0.310    0.071    4.370    0.000
  f2 =~
    x2                0.508    0.055    9.296    0.000
    x3                0.555    0.062    9.000    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 ~~
    x4                0.318    0.084    3.781    0.000
  f2 ~~
    x4                0.365    0.053    6.899    0.000
  f1 ~~
    f2                0.681    0.143    4.775    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.380    0.051    7.447    0.000
   .x5                0.332    0.048    6.901    0.000
   .x2                0.211    0.042    5.031    0.000
   .x3                0.300    0.053    5.689    0.000
    x4                0.454    0.045   10.000    0.000
    f1                1.000
    f2                1.000
```



ค่าน้ำหนักองค์ประกอบ สหสัมพันธ์ระหว่างองค์ประกอบ และขนาดความแปรปรวนของค่าคงเหลือ เท่ากับโมเดลเดิม

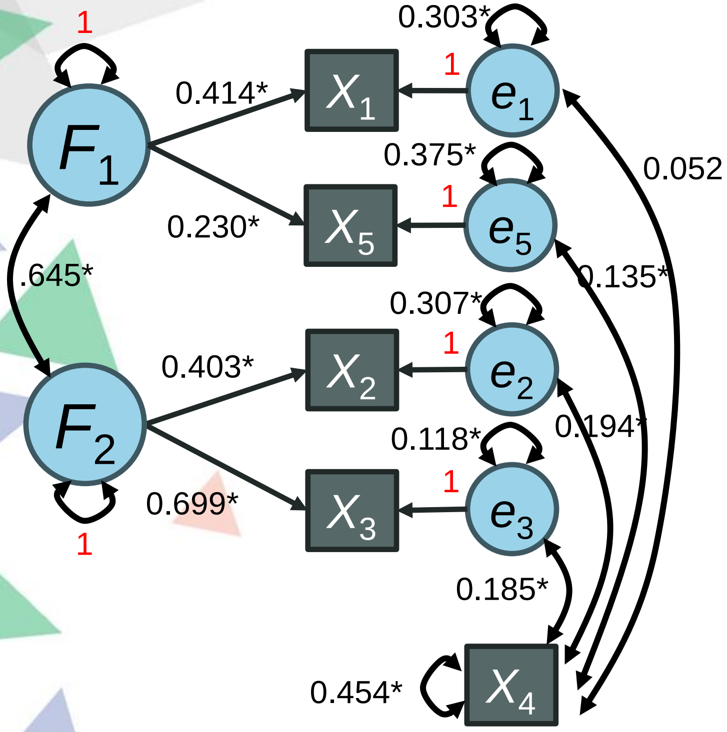
โมเดลนี้ เอา X_4 มาหาความสัมพันธ์กับ e_1, e_2, e_3 และ e_5

โมเดลนี้เป็นการเอา X_4 เข้ามาโดยไม่ยุ่งกับโมเดลเดิม
โมเดลนี้มักใช้ในการประมาณค่าสูญหาย โดยใช้ X_4 เป็นตัวแปรช่วย (Auxiliary Variables)

$\boldsymbol{\Theta}$ ไม่ Positive Definite

```
> msingle2 <- '
+ f1 =~ x1 + x5
+ f2 =~ x2 + x3
+ x4 ~~ x1 + x5 + x2 + x3
+ '
> outsingle2 <- cfa(msingle2, data=dat, std.lv=TRUE)
Warning message:
In lav_object_post_check(object) :
  lavaan WARNING: the covariance matrix of the residuals of the observed
                  variables (theta) is not positive definite;
                  use lavInspect(fit, "theta") to investigate.
> summary(outsingle2)
lavaan 0.6-12 ended normally after 27 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        14

  Number of observations                           200

Model Test User Model:

  Test statistic                                 0.851
  Degrees of freedom                                 1
  P-value (Chi-square)                           0.356

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    x1                0.414    0.091    4.553    0.000
    x5                0.230    0.063    3.653    0.000
  f2 =~
    x2                0.403    0.064    6.279    0.000
    x3                0.699    0.091    7.646    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
 .x1 ~~
    x4                0.052    0.033    1.577    0.115
 .x5 ~~
    x4                0.135    0.032    4.218    0.000
 .x2 ~~
    x4                0.194    0.035    5.567    0.000
 .x3 ~~
    x4                0.185    0.039    4.712    0.000
  f1 ~~
    f2                0.645    0.145    4.438    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.303    0.073    4.167    0.000
   .x5                0.375    0.043    8.795    0.000
   .x2                0.307    0.048    6.334    0.000
   .x3                0.118    0.114    1.041    0.298
    x4                0.451    0.045   10.026    0.000
    f1                1.000
    f2                1.000
```

Chi-square, *df* ไม่เปลี่ยน



หา $\mathbf{\Theta}$

```
> lavInspect(outsingle2, "theta")
   x1    x5    x2    x3    x4
x1 0.303
x5 0.000 0.375
x2 0.000 0.000 0.307
x3 0.000 0.000 0.000 0.118
x4 0.052 0.135 0.194 0.185 0.451
> eigen(lavInspect(outsingle2, "theta"))
eigen() decomposition
$values
[1]  0.679976503  0.357445823  0.302906146  0.218673320 -0.004401651

$vectors
           [,1]        [,2]         [,3]       [,4]        [,5]
[1,] -0.1086322  0.10835779  0.965535461  0.1916636 -0.08639426
[2,] -0.3505744 -0.87855250 -0.010389565  0.2681835 -0.18224380
[3,] -0.4120169  0.44195614 -0.259944986  0.6822471 -0.31919318
[4,] -0.2608867  0.08885674  0.005572471 -0.5738039 -0.77120724
[5,] -0.7921343  0.11481876  0.005557127 -0.3108546  0.51252154
```

หา eigenvalues, eigenvectors ของ $\mathbf{\Theta}$

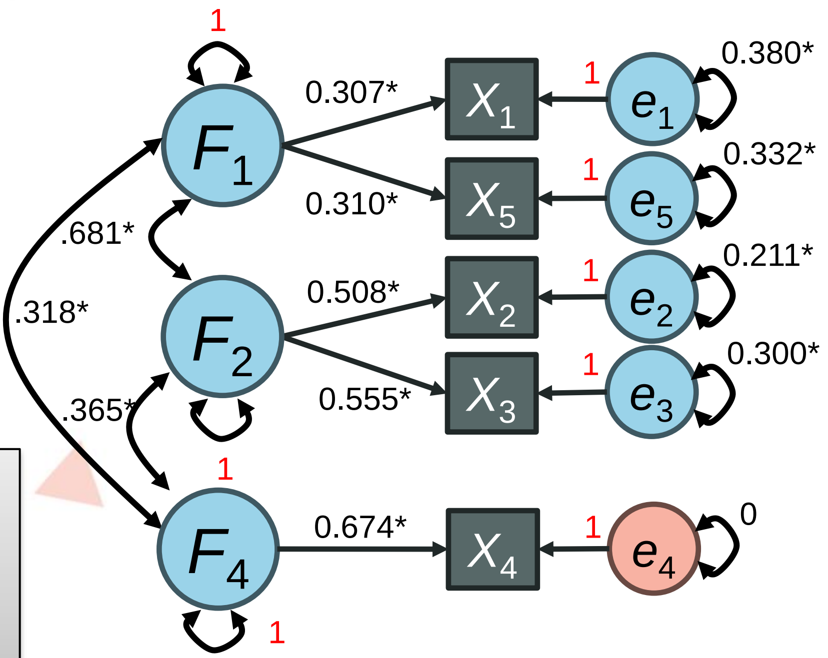
ไม่ Positive Definite หมายความว่า เมทริกซ์ความแปรปรวนร่วม (Covariance Matrix) นั้น เป็นไปไม่ได้ตามความเป็นจริง

องค์ประกอบที่มีตัวบ่งชี้เดียว

- Positive Definite สามารถตรวจสอบได้โดยการทำ Eigendecomposition เป็นการทำ Principal Component Analysis (PCA) ของตัวแปรทั้งหมด หากพบว่า Component ใดมีความแปรปรวนเป็นลบ (Eigenvalue เป็นลบ) หมายความว่าการแปลงเป็น Components ทำไม่ได้จริง
- ตัวอย่างของเมทริกซ์ที่เป็นไปไม่ได้ เช่น x1 สัมพันธ์กับ x2 = .9 และ x1 สัมพันธ์กับ x3 = .8 แต่ x2 ไปสัมพันธ์กับ x3 = -.5 เมื่อ x1, x2, และ x3 ไปในทิศทางเดียวกันแล้ว x2 และ x3 จะไปสัมพันธ์กันแบบติดลบได้อย่างไร
  - เมทริกซ์สหสัมพันธ์นี้ ก็จะทำให้ความแปรปรวนของ Component ติดลบ
- โดยปกติค่าความแปรปรวน Component แทบเป็นไปไม่ได้ หากวิเคราะห์ข้อมูลจริง

องค์ประกอบที่มีตัวบ่งชี้เดียว

- สาเหตุที่ความแปรปรวนของ Component ติดลบได้ เนื่องจาก SEM ประมาณค่าพารามิเตอร์แต่ละตัวแยกจากกัน ไม่ได้ประมาณค่าพร้อมกัน
- พอได้ผลสุดท้ายจาก ML มาตรวจสอบดันไปเจอว่า อ้าว ค่านี้มันเป็นไปไม่ได้ในความเป็นจริง!!! จึงขึ้น Warning ขึ้นมา
- นักวิเคราะห์ต้องพิจารณาว่าควรปรับโมเดล หรือค่า Warning นี้พอรับได้ ในกรณีนี้พอรับได้ เนื่องจาก X_4 ไม่ได้ใช้อะไรในโมเดลอยู่แล้ว แค่นำมาช่วยจัดการค่าสูญหายในบทบาทของตัวแปรช่วย (Auxiliary Variables)



ใช้วิธีกำหนดสเกลโดยให้ความแปรปรวนของ F_4 เท่ากับ 1 อาจเปลี่ยนไปกำหนดน้ำหนักองค์ประกอบเท่ากับ 1 แทนได้

โมเดลนี้ เอา X_4 มาหาความสัมพันธ์กับ F_1 และ F_2

บางโปรแกรม ไม่สามารถเอา X_4 มาหาความสัมพันธ์กับ F_1 และ F_2 ได้โดยตรง จึงต้องสร้าง F_4 ขึ้นมา แล้วกำหนดให้ความแปรปรวนของ e_4 เท่ากับ 0

```
> msingle1a <- '
+ f1 =~ x1 + x5
+ f2 =~ x2 + x3
+ f4 =~ x4
+ x4 ~~ 0*x4
+ '
> outsingle1a <- cfa(msingle1a, data=dat, std.lv=TRUE)
> summary(outsingle1a)
lavaan 0.6-12 ended normally after 30 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        12

  Number of observations                           200

Model Test User Model:

  Test statistic                                19.066
  Degrees of freedom                                 3
  P-value (Chi-square)                           0.000

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 =~
    x1                0.307    0.072    4.252    0.000
    x5                0.310    0.071    4.370    0.000
  f2 =~
    x2                0.508    0.055    9.296    0.000
    x3                0.555    0.062    9.000    0.000
  f4 =~
    x4                0.674    0.034   20.000    0.000
```

.674 เท่ากับรากที่ 2 ของความแปรปรวนของ X_4

```
Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  f1 ~~
    f2                0.681    0.143    4.775    0.000
    f4                0.472    0.119    3.967    0.000
  f2 ~~
    f4                0.541    0.066    8.172    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x4                0.000
   .x1                0.380    0.051    7.447    0.000
   .x5                0.332    0.048    6.901    0.000
   .x2                0.211    0.042    5.031    0.000
   .x3                0.300    0.053    5.689    0.000
    f1                1.000
    f2                1.000
    f4                1.000
```