โมเดลเส้นโค้งพัฒนาการแบบซับซ้อน (Complex Growth Curve Model)

โมเดลพหุระดับ (Multilevel Modeling)

สันทัด พรประเสริฐมานิต

โครงร่างการนำเสนอ

- รูปแบบการเปลี่ยนแปลง
- โครงสร้างความสัมพันธ์ค่าคงเหลือ (Residual Correlation)
- สัมประสิทธิ์การทำนาย

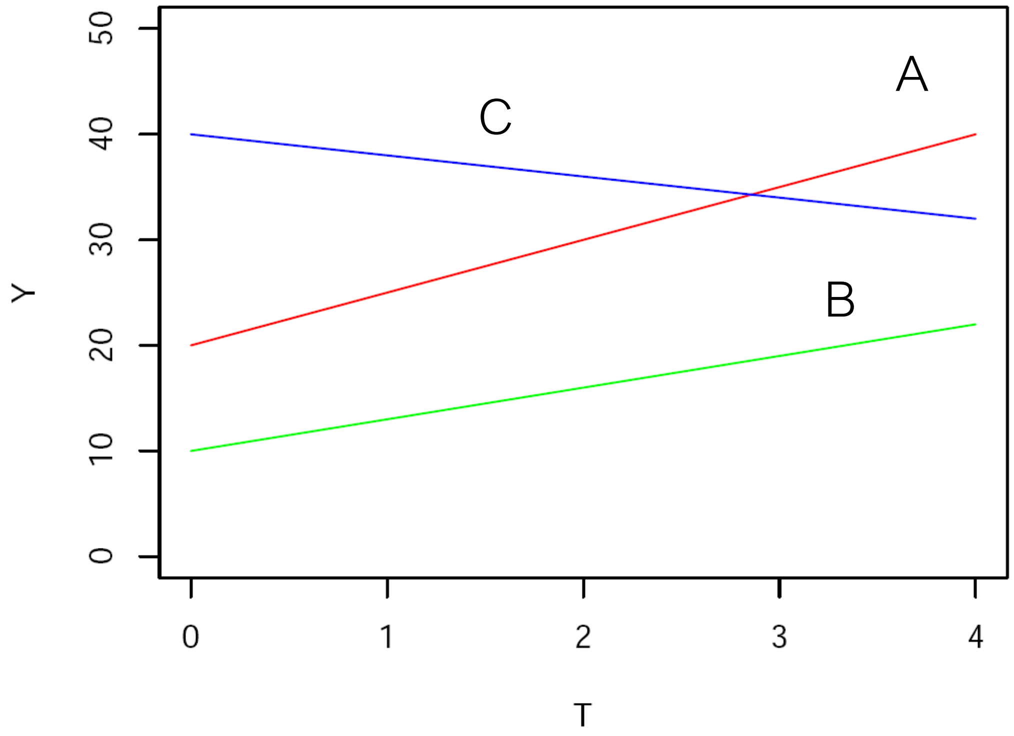
รูปแบบการเปลี่ยนแปลง

โมเดลการเปลี่ยนแปลงเชิงเส้น

$$Y = b_0 + b_1 t$$

$$A: \ Y = 20 + 5t$$
$$B: \ Y = 10 + 3t$$
$$C: \ Y = 40 - 2t$$

t	*Y* (A)	*Y* (B)	*Y* (C)
0	20	10	40
1	25	13	38
2	30	16	36
3	35	19	34
4	40	22	32



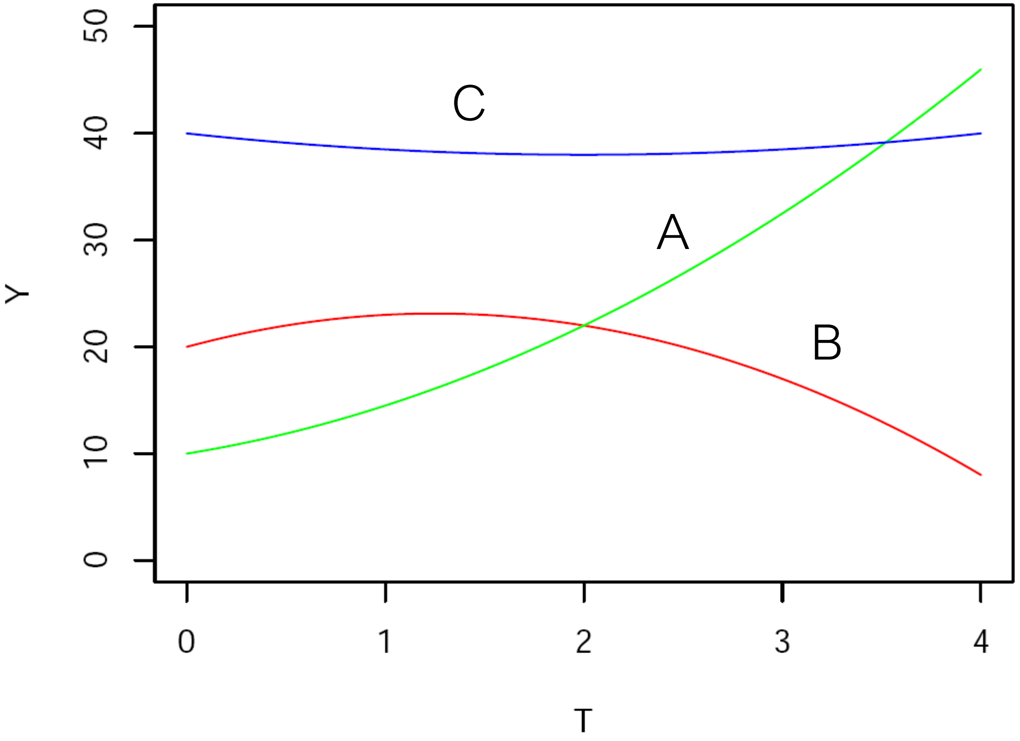
รูปแบบการเปลี่ยนแปลง

โมเดลการเปลี่ยนแปลงแบบสมการยกกำลังสอง (Quadratic curve)

$$Y = b_0 + b_1 t + b_2 t^2$$

$$A: Y = 20 + 5t - 2t^2$$
$$B: Y = 10 + 3t + 1.5t^2$$
$$C: Y = 40 - 2t - 0.5t^2$$

t	*Y* (A)	*Y* (B)	*Y* (C)
0	20	10	40
1	23	14.5	38.5
2	22	22	38
3	17	32.5	38.5
4	8	46	40



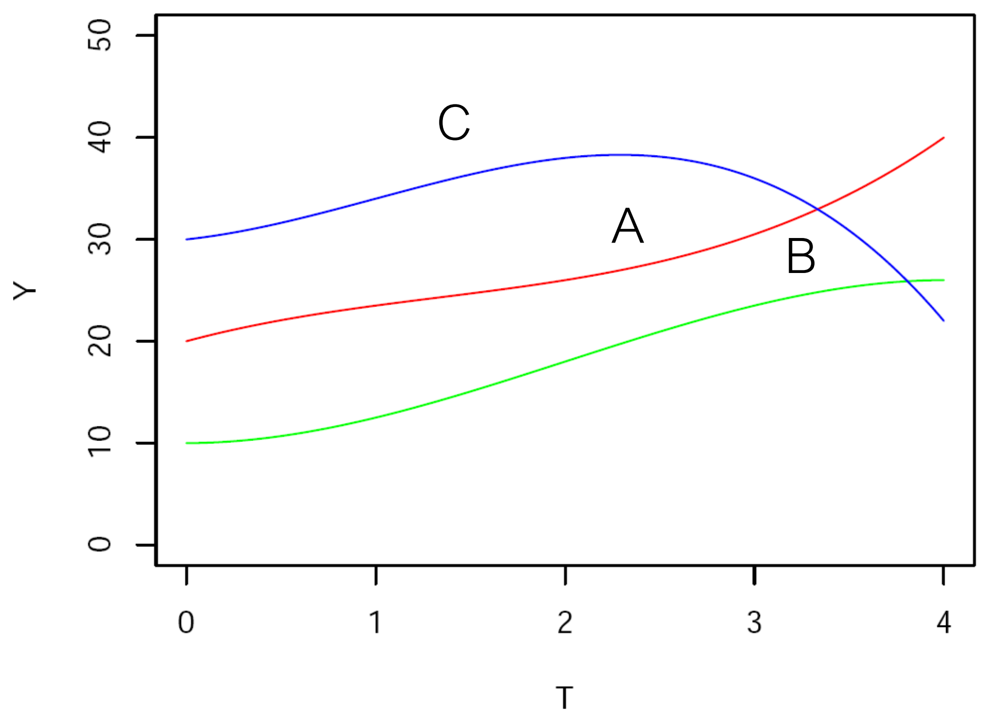
รูปแบบการเปลี่ยนแปลง

โมเดลการเปลี่ยนแปลงแบบสมการยกกำลังสาม (Cubic curve)

$$Y = b_0 + b_1 t + b_2 t^2 + b_3 t^3$$

$$A: Y = 20 + 5t - 2t^2 + 0.5t^3$$
$$B: Y = 10 + 3t^2 - 0.5t^3$$
$$C: Y = 30 + 2t + 3t^2 - t^3$$

t	*Y* (A)	*Y* (B)	*Y* (C)
0	20	10	30
1	23.5	12.5	34
2	26	18	38
3	30.5	23.5	36
4	40	26	22



รูปแบบการเปลี่ยนแปลง

โมเดลการเปลี่ยนแปลงแบ่งช่วงย่อย (Spline Model)

t	t_1	t_2
0	0	0
1	1	0
2	2	0
3	2	1
4	2	2

แบ่งเวลาออกเป็น 2 ช่วง โดยให้ $t = t_1 + t_2$

แบ่งเวลาออกเป็น 3 ช่วง โดยให้ $t = t_1 + t_2 + t_3$

จากสมการดังกล่าว สามารถแบ่งการเปลี่ยนแปลงเชิงเส้นเป็นดังนี้

$$Y = b_0 + b_1 t = b_0 + b_1 t_1 + b_1 t_2$$

สมการนี้สามารถให้สัมประสิทธิ์ถดถอยของ t_1 และ t_2 แตกต่างกันได้

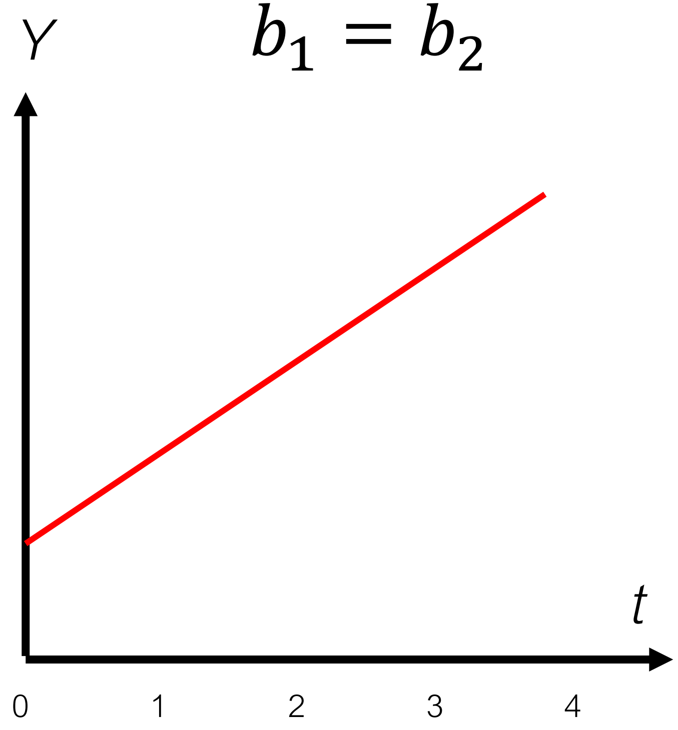
$$Y = b_0 + b_1 t_1 + b_2 t_2$$

รูปแบบการเปลี่ยนแปลง

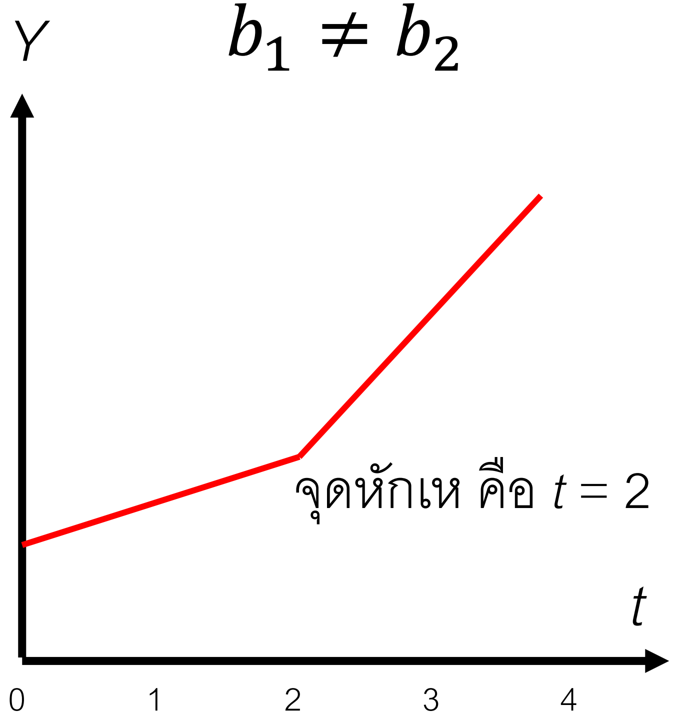
$$Y = b_0 + b_1 t_1 + b_2 t_2$$

t	t_1	t_2
0	0	0
1	1	0
2	2	0
3	2	1
4	2	2

$b_1 = b_2$



$b_1 \neq b_2$

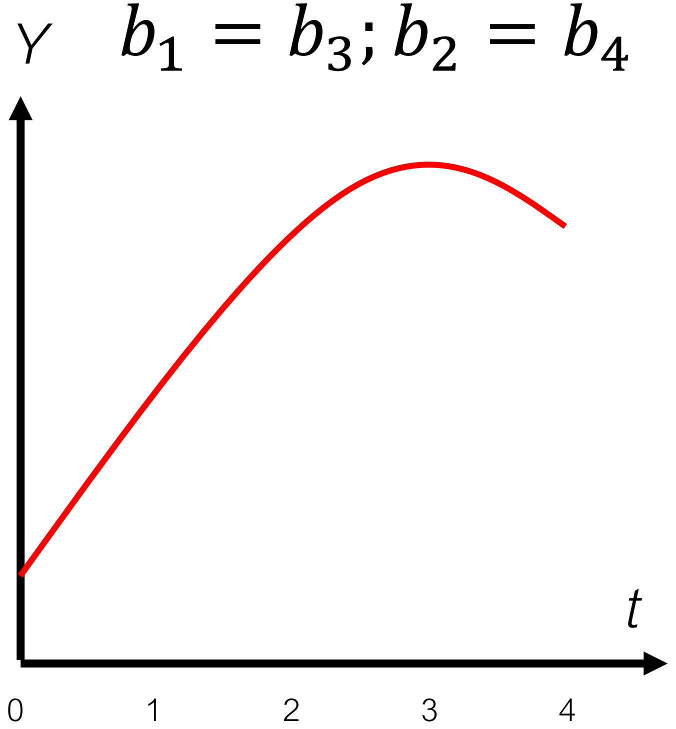


รูปแบบการเปลี่ยนแปลง

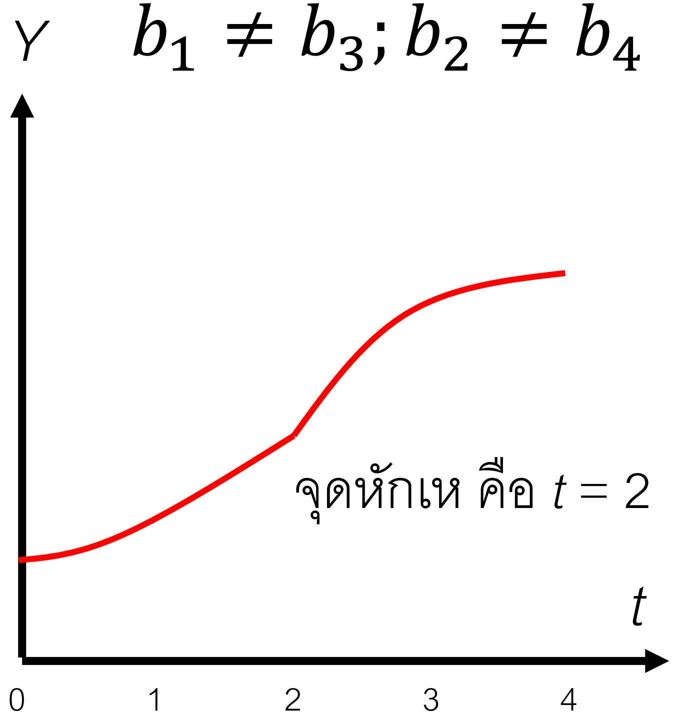
$$Y = b_0 + b_1 t_1 + b_2 t_1^2 + b_3 t_2 + b_4 t_2^2$$

t	t_1	t_2
0	0	0
1	1	0
2	2	0
3	2	1
4	2	2

$b_1 = b_3; b_2 = b_4$



$b_1 \neq b_3; b_2 \neq b_4$



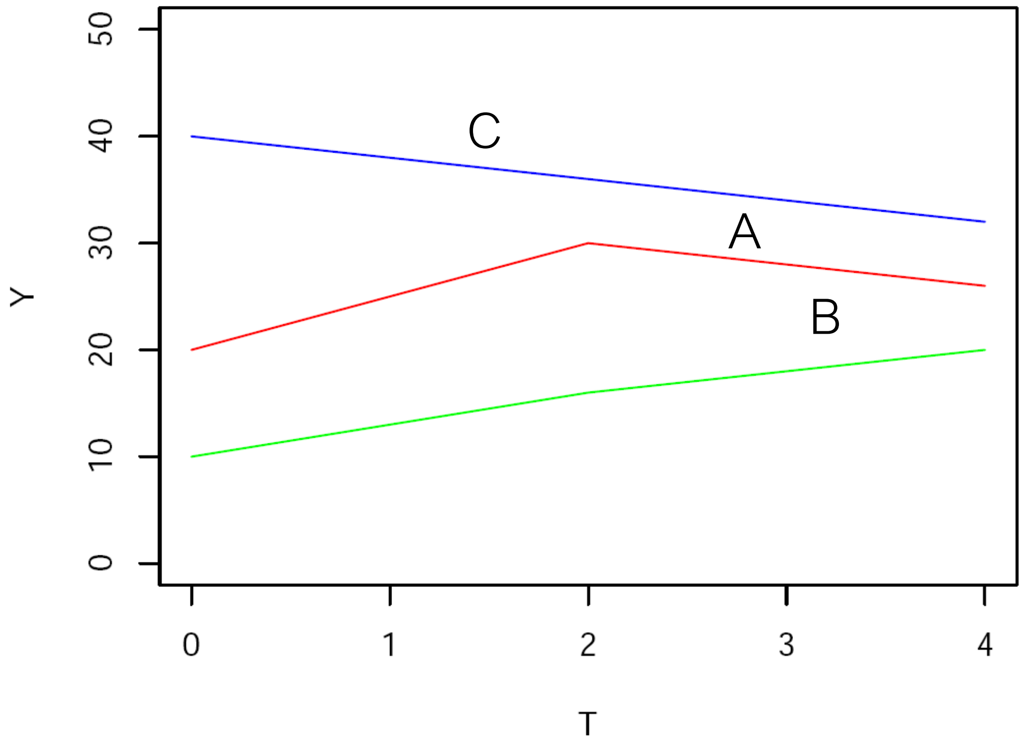
รูปแบบการเปลี่ยนแปลง

โมเดลการเปลี่ยนแปลงแบ่งช่วงย่อยเชิงเส้น (Linear Spline Model)

$$Y = b_0 + b_1 t + b_2 t$$

$$A: Y = 20 + 5t_1 - 2t_2$$
$$B: Y = 10 + 3t_1 + 2t_2$$
$$C: Y = 40 - 2t_1 - 2t_2$$

t	t_1	t_2	Y (A)	Y (B)	Y (C)
0	0	0	20	10	40
1	1	0	25	13	38
2	2	0	30	16	36
3	2	1	28	18	34
4	2	2	26	20	32



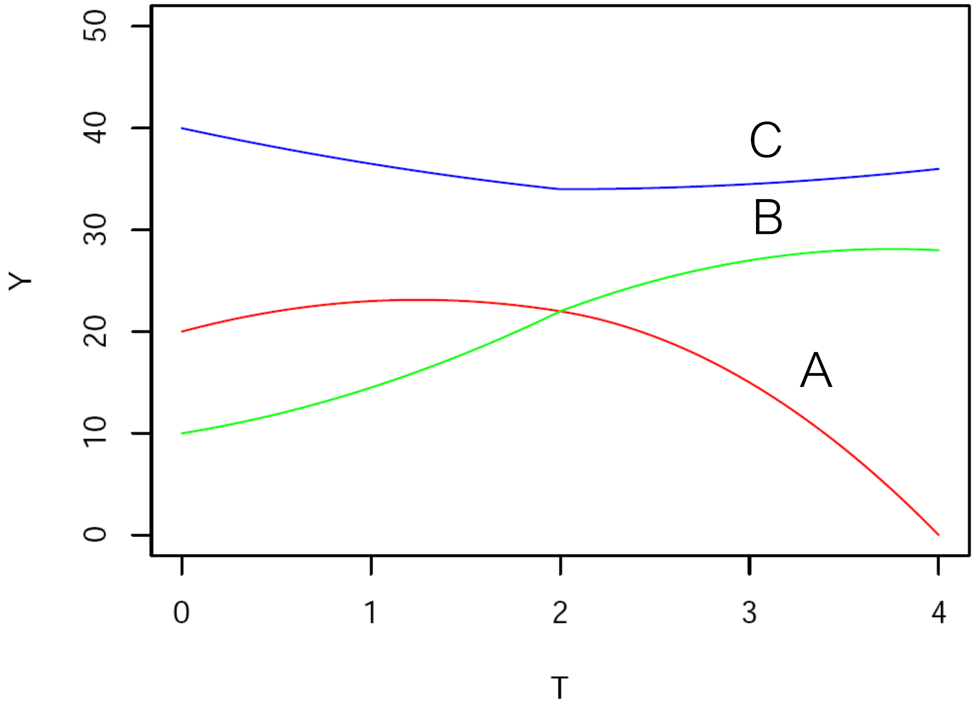
รูปแบบการเปลี่ยนแปลง

โมเดลการเปลี่ยนแปลงแบ่งช่วงย่อยกำลังสอง (Quadratic Spline Model)

$$Y = b_0 + b_1 t + b_2 t_2^2 + b_3 t_2 + b_4 t_2^2$$

$$A\colon\ Y = 20 + 5t_1 - 2t_1^2 - 3t_2 - 4t_2^2$$
$$B\colon\ Y = 10 + 3t_1 + 1.5t_1^2 + 7t_2 - 2t_2^2$$
$$C\colon\ Y = 40 - 4t_1 - 0.5t_1^2 - 2t_2^2$$

t	t_1	t_2	*Y* (A)	*Y* (B)	*Y* (C)
0	0	0	20	10	40
1	1	0	23	14.5	36.5
2	2	0	22	22	34
3	2	1	15	27	34.5
4	2	2	0	28	36



โมเดลการเปลี่ยนแปลงแบ่งช่วงย่อยกำลังสอง (Quadratic Spline Model)

$$Y = b_0 + b_1 t + b_2 t_2^2 + b_3 t_2 + b_4 t_2^2 + b_5 t_3 + b_6 t_3^2$$

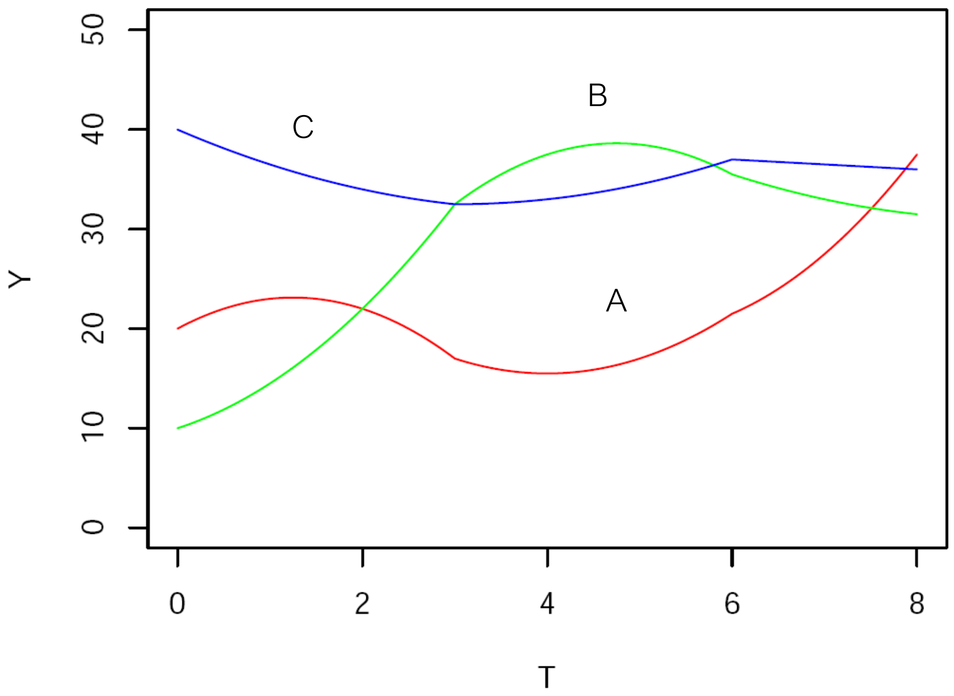
แบ่งสามช่วง

$$A: Y = 20 + 5t_1 - 2t_1^2 - 3t_2 + 1.5t_2^2 + 4t_3 + 2t_3^2$$

$$B: Y = 10 + 3t_1 + 1.5t_1^2 + 7t_2 - 2t_2^2 - 3t_3 + 0.5t_3^2$$

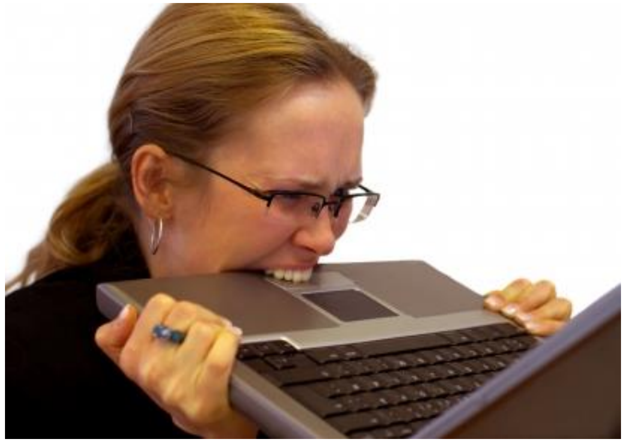
$$C: Y = 40 - 4t_1 - 0.5t_1^2 - 2t_2^2 - 0.5t_3$$

t	t_1	t_2	t_3	Y (A)	Y (B)	Y (C)
0	0	0	0	20	10	40
1	1	0	0	23	14.5	36.5
2	2	0	0	22	22	34
3	3	0	0	17	32.5	32.5
4	3	1	0	15.5	37.5	33
5	3	2	0	17	38.5	34.5
6	3	3	0	21.5	35.5	37
7	3	3	1	27.5	33	36.5
8	3	3	2	37.5	31.5	36



รูปแบบการเปลี่ยนแปลง

นักวิจัยต้องการตรวจสอบผลการบำบัดความเครียดจากสถาบันแห่งหนึ่ง

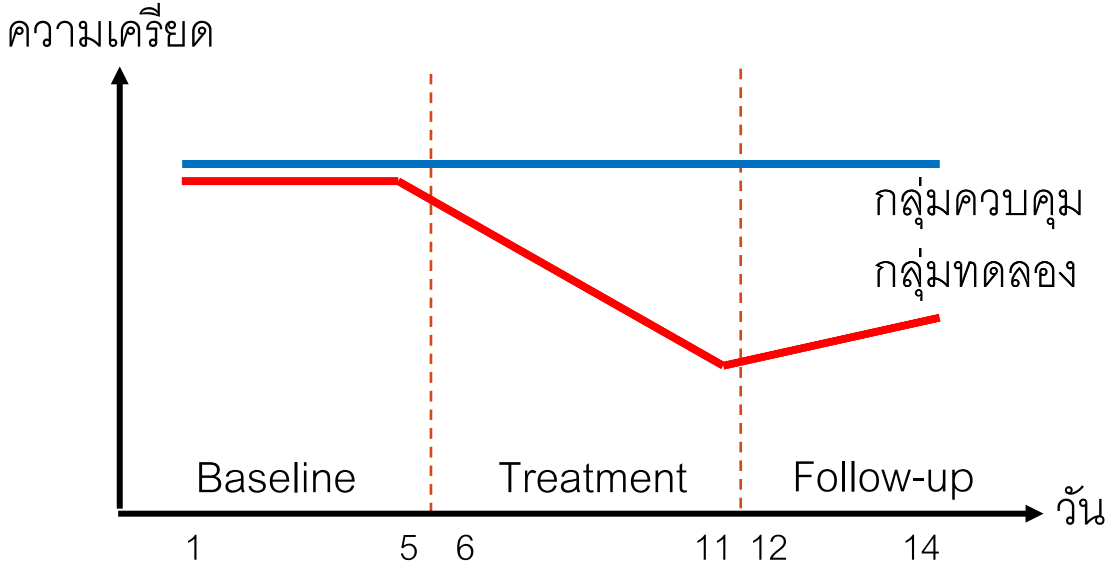


นักวิจัยเก็บข้อมูลจากคนไข้ 1000 คน โดยแบ่งคนไข้เป็น 2 กลุ่ม คือ กลุ่มบำบัดปกติ และกลุ่มควบคุม โดยเก็บข้อมูลต่อเนื่อง 14 วัน

ระยะที่ 1 : วัดข้อมูลพื้นฐาน (วันที่ 1-5)

ระยะที่ 2 : บำบัดคนไข้ (วันที่ 6-11)

ระยะที่ 3 : ติดตามผล (วันที่ 12-14)



เวลามีทั้งหมด 14 วัน จะใส่ค่าตั้งแต่ 0 – 13 โดยค่า 0 คือค่าความเครียดก่อนเข้ารับบำบัด

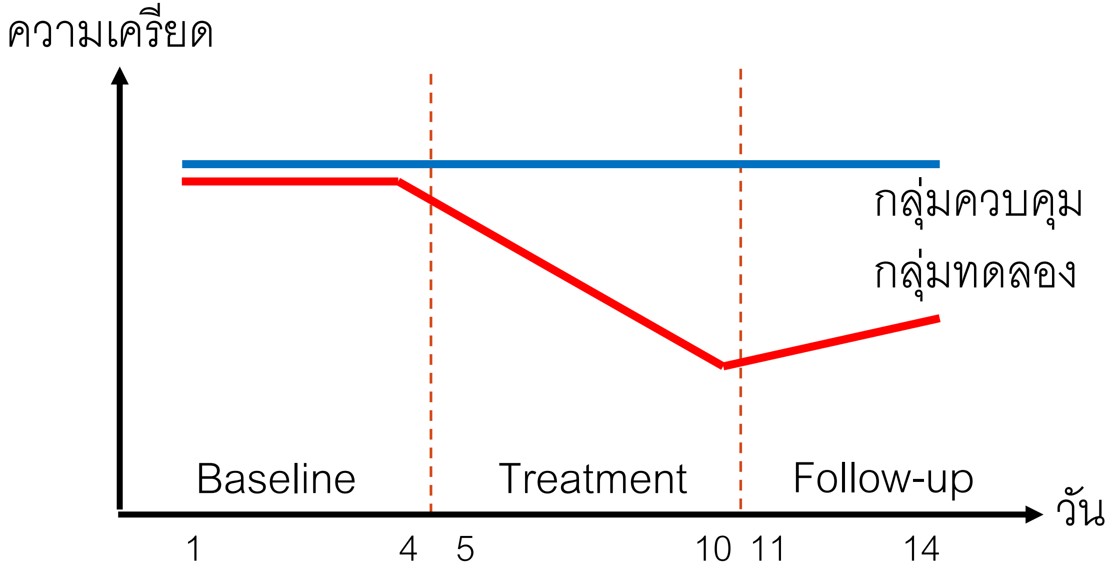
t	t_c	t_1	t_2	t_3
1	0	0	0	0
2	1	1	0	0
3	2	2	0	0
4	3	3	0	0
5	4	4	0	0
6	5	4	1	0
7	6	4	2	0
8	7	4	3	0
9	8	4	4	0
10	9	4	5	0
11	10	4	6	0
12	11	4	6	1
13	12	4	6	2
14	13	4	6	3

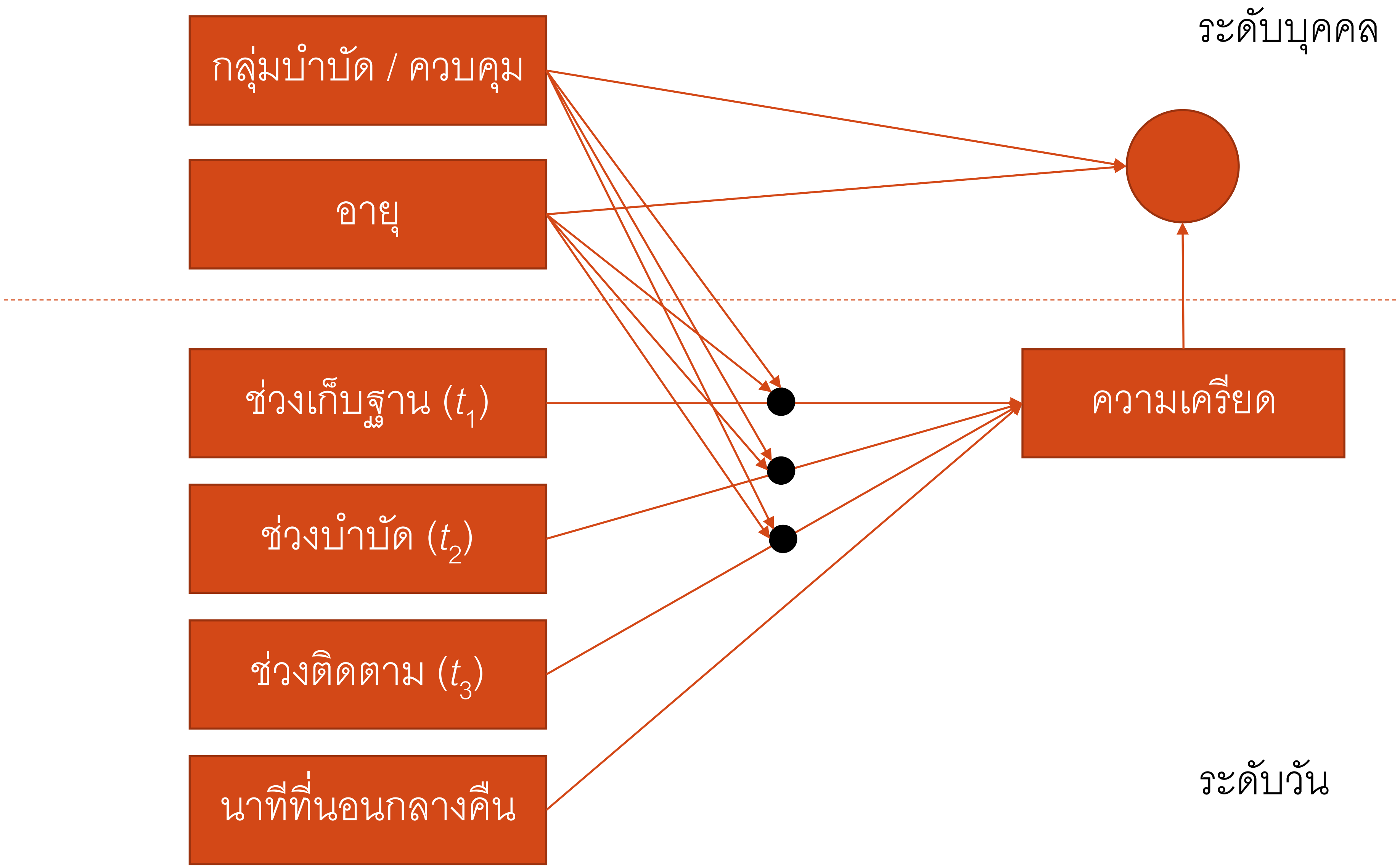
$$t_c = t_1 + t_2 + t_3$$

t_1 = การเปลี่ยนแปลงเชิงเส้นช่วง Baseline

t_2 = การเปลี่ยนแปลงเชิงเส้นช่วง Treatment

t_3 = การเปลี่ยนแปลงเชิงเส้นช่วง Follow-up





ระดับบุคคล
กลุ่มบำบัด / ควบคุม
อายุ
ช่วงเก็บฐาน (t_1)
ช่วงบำบัด (t_2)
ช่วงติดตาม (t_3)
นาทีที่นอนกลางคืน
ความเครียด
ระดับวัน

รูปแบบการเปลี่ยนแปลง

```
> dat1 <-read.table("lecture9ex1.csv", sep=",", header=TRUE)
> psych::describe(dat1)
       vars     n    mean      sd median trimmed     mad min   max range  skew kurtosis    se
rowid     1 14000 7000.50 4041.60 7000.5 7000.50 5189.10   1 14000 13999  0.00    -1.20 34.16
pid       2 14000  500.50  288.69  500.5  500.50  370.65   1  1000   999  0.00    -1.20  2.44
time      3 14000    7.50    4.03    7.5    7.50    5.19   1    14    13  0.00    -1.21  0.03
stress    4 14000   50.21    9.85   51.0   50.39   10.38  11    83    72 -0.20     0.10  0.08
treat     5 14000    0.50    0.50    0.5    0.50    0.74   0     1     1  0.00    -2.00  0.00
sleep     6 14000  423.58   58.69  423.0  423.53   60.79 226   635   409  0.01    -0.13  0.50
age       7 14000   36.64    6.20   36.0   36.58    5.93  18    60    42  0.11     0.00  0.05
```

- ตัวแปรเวลา (time) ที่แสดงถึงวันที่วัด
- แบ่งการเปลี่ยนแปลงเชิงเส้นออกเป็น 3 ส่วน คือ ช่วงวัดข้อมูลพื้นฐาน, ช่วงบำบัด, และช่วงติดตามผล

```
> dat1$timec <- dat1$time - 1
> dat1$timebase <- dat1$timec
> dat1[dat1$timebase > 4, "timebase"] <- 4
> dat1$timetreat <- dat1$timec - dat1$timebase
> dat1[dat1$timetreat > 6, "timetreat"] <- 6
> dat1$timefollow <- dat1$timec - dat1$timebase - dat1$timetreat
```

รูปแบบการเปลี่ยนแปลง

```
> dat1 <-read.table("lecture9ex1.csv", sep=",", header=TRUE)
> psych::describe(dat1)
       vars     n    mean      sd median trimmed     mad min   max range  skew kurtosis    se
rowid     1 14000 7000.50 4041.60 7000.5 7000.50 5189.10   1 14000 13999  0.00    -1.20 34.16
pid       2 14000  500.50  288.69  500.5  500.50  370.65   1  1000   999  0.00    -1.20  2.44
time      3 14000    7.50    4.03    7.5    7.50    5.19   1    14    13  0.00    -1.21  0.03
stress    4 14000   50.21    9.85   51.0   50.39   10.38  11    83    72 -0.20     0.10  0.08
treat     5 14000    0.50    0.50    0.5    0.50    0.74   0     1     1  0.00    -2.00  0.00
sleep     6 14000  423.58   58.69  423.0  423.53   60.79 226   635   409  0.01    -0.13  0.50
age       7 14000   36.64    6.20   36.0   36.58    5.93  18    60    42  0.11     0.00  0.05
```

- ตัวแปรควบคุมที่เปลี่ยนแปลงตามเวลา (Time-varying covariate) คือตัวแปรแทรกซ้อนที่อยู่ในระดับการวัดซ้ำ
- ตัวแปรควบคุมที่ไม่เปลี่ยนแปลงตามเวลา (Time-invariant covariate) คือตัวแปรแทรกซ้อนที่อยู่ในระดับบุคคล

ตัวแปรทั้งสองระดับ อาจมีปฏิสัมพันธ์กับตัวแปรเวลาได้

โมเดลว่าง

```
> out1a <- lmer(stress ~ 1 + (1|pid), data=dat1, REML=FALSE)
> summary(out1a)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: stress ~ 1 + (1 | pid)
   Data: dat1

     AIC      BIC   logLik deviance df.resid
103315.8 103338.4 -51654.9 103309.8    13997

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7791 -0.6489  0.0212  0.6755  3.2755

Random effects:
 Groups   Name        Variance Std.Dev.
 pid      (Intercept)  9.076   3.013
 Residual             88.021   9.382
Number of obs: 14000, groups:  pid, 1000

Fixed effects:
            Estimate Std. Error t value
(Intercept)  50.2054     0.1239     405
```

ความแปรปรวนระดับบุคคล = 9.08

ความแปรปรวนระดับการวัดซ้ำภายในบุคคล = 88.02

ICC = 9.08 / (9.08 + 88.02) = .093

ค่าเฉลี่ยของความเครียดระหว่างทุกคนทุกวัน = 50.20 คะแนน

ใส่ตัวแปรเวลา โดยอัตราการเปลี่ยนแปลงเชิงเส้นตรง ยังไม่แบ่งช่วง ยังไม่เปลี่ยนตามบุคคล

```
> out1b <- lmer(stress ~ 1 + timec + (1|pid), data=dat1, REML=FALSE)
> summary(out1b)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: stress ~ 1 + timec + (1 | pid)
   Data: dat1

     AIC      BIC   logLik deviance df.resid
102801.3 102831.5 -51396.6 102793.3    13996

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8198 -0.6463  0.0086  0.6626  3.2683

Random effects:
 Groups   Name        Variance Std.Dev.
 pid      (Intercept)  9.321   3.053
 Residual             84.592   9.197
Number of obs: 14000, groups:  pid, 1000

Fixed effects:
            Estimate Std. Error t value
(Intercept) 53.08240    0.17628  301.13
timec       -0.44262    0.01928  -22.95
```

ความแปรปรวนระดับบุคคล = 9.32

ความแปรปรวนระดับการวัดซ้ำภายในบุคคลคงเหลือ = 84.59

ค่าเฉลี่ยของความเครียดในวันที่ 1 (ก่อนบำบัด) = 53.08 คะแนน

เมื่อผ่านไปแต่วัน ความเครียดลดลง วันละ 0.44 คะแนน ซึ่งลดลงอย่างมีนัยสำคัญ

ใส่ตัวแปรเวลา เปลี่ยนแปลงเชิงเส้นตรงแบบแบ่งช่วง ยังไม่เปลี่ยนตามบุคคล

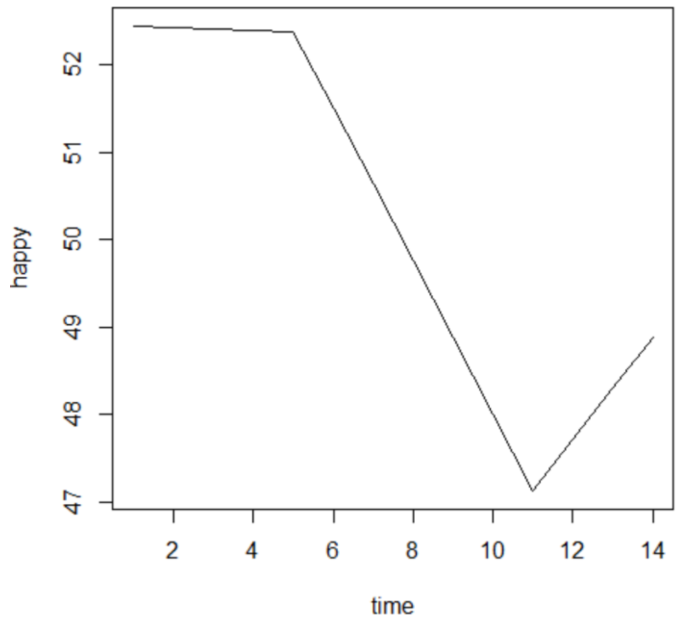
```
> out1c <- lmer(stress ~ 1 + timebase + timetreat
+              + timefollow + (1|pid), data=dat1, REML=FALSE)
> summary(out1c)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: stress ~ 1 + timebase + timetreat + timefollow + (1 | pid)
   Data: dat1

     AIC      BIC   logLik deviance df.resid
102693.1 102738.3 -51340.5 102681.1    13994

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8092 -0.6436  0.0062  0.6596  3.3402

Random effects:
 Groups   Name        Variance Std.Dev.
 pid      (Intercept)  9.373   3.062
 Residual             83.865   9.158
Number of obs: 14000, groups:  pid, 1000

Fixed effects:
            Estimate Std. Error t value
(Intercept) 52.44186    0.23946 218.998
timebase    -0.01506    0.07770  -0.194
timetreat   -0.87487    0.04711 -18.572
timefollow   0.58362    0.10854   5.377
```



ค่าเฉลี่ยของความเครียดในวันที่ 1 (ก่อนบำบัด) = 52.44 คะแนน

ความเครียดช่วงเก็บฐาน ลดลงวันละ 0.02 คะแนน ซึ่งไม่ถึงระดับนัยสำคัญ

ความเครียดช่วงทดลอง ลดลงวันละ 0.87 คะแนน ซึ่งถึงระดับนัยสำคัญ

ความเครียดช่วงติดตาม เพิ่มขึ้นวันละ 0.58 คะแนน ซึ่งถึงระดับนัยสำคัญ

เปรียบเทียบระหว่างโมเดลการเปลี่ยนแปลงเชิงเส้น แบบแบ่งและไม่แบ่งช่วงเวลา

```
> anova(out1b, out1c)
Data: dat1
Models:
out1b: stress ~ 1 + timec + (1 | pid)
out1c: stress ~ 1 + timebase + timetreat + timefollow + (1 | pid)
      Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
out1b  4 102801 102831 -51397   102793
out1c  6 102693 102738 -51341   102681 112.21      2  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ความแตกต่างถึงระดับนัยสำคัญ แสดงว่าการแบ่งช่วงเวลาเป็นสิ่งที่ควรทำ

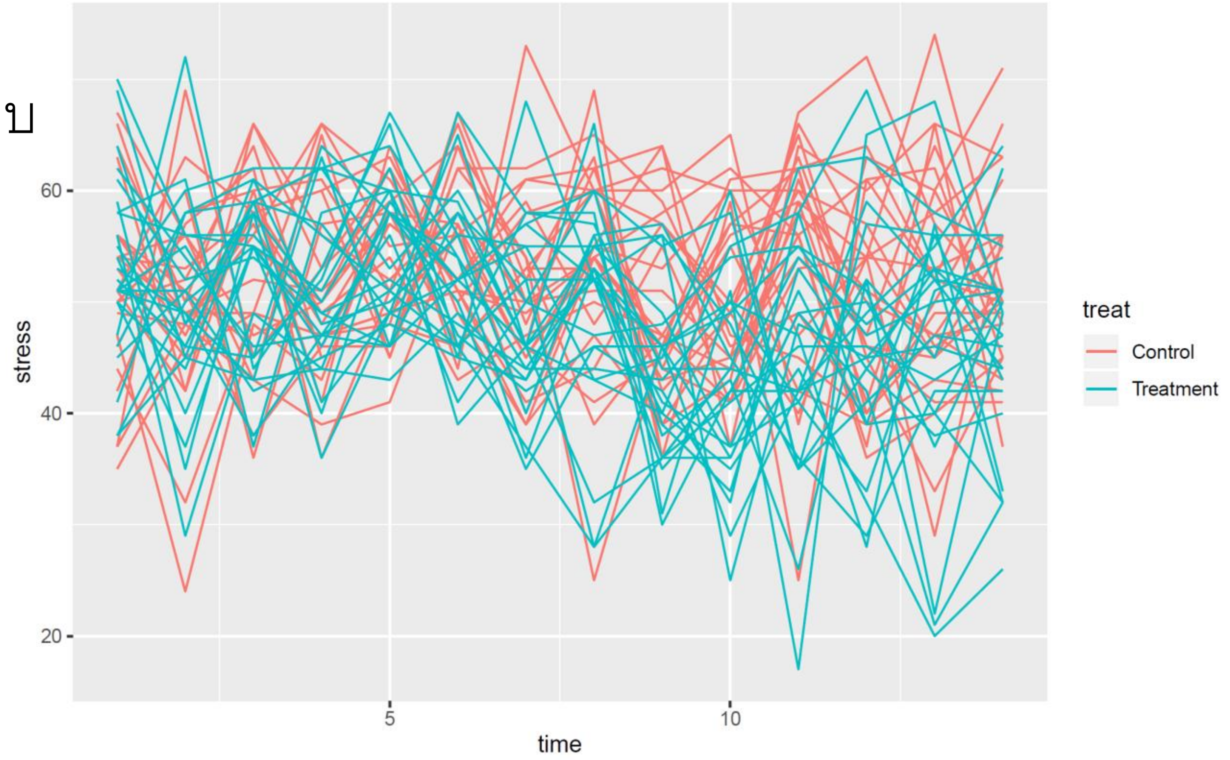
การเปรียบเทียบ 2 โมเดลนี้ เหมือนการตรวจสอบว่า การจำกัดให้การเปลี่ยนแปลงทั้งสามช่วงเท่ากันนั้น เหมาะสมหรือไม่ ซึ่งผลออกมาว่าไม่เหมาะสม

$$Y_{ij} = \gamma_{00} + \gamma_{10}T_{1ij} + \gamma_{20}T_{2ij} + \gamma_{30}T_{3ij} + u_{0j} + e_{ij}$$

$$H_0: \gamma_{10} = \gamma_{20} = \gamma_{30}$$

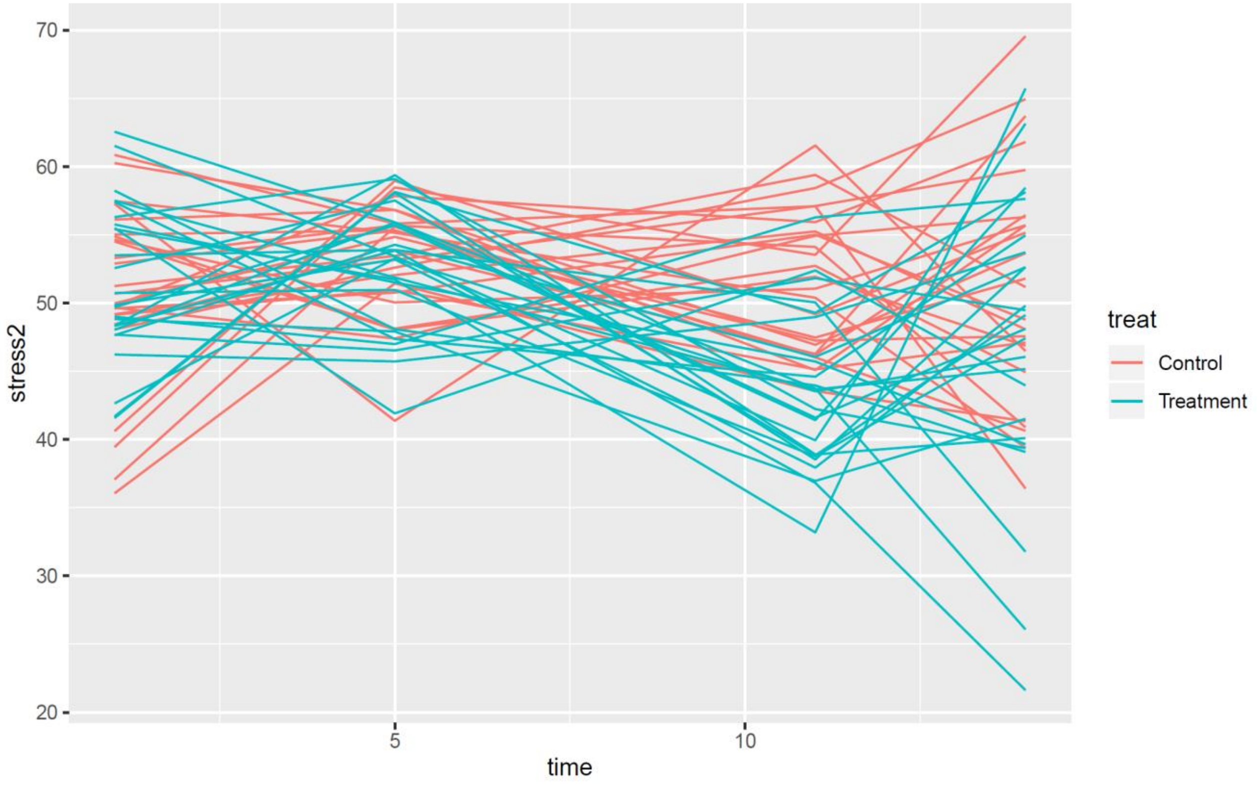
```
> datl <- split(dat1, dat1$pid)
> stresspred <- lapply(datl, function(x) { predict(lm(stress ~ timebase + timetreat + timefollow, data=x)) })
> dat1$stress2 <- do.call(c, stresspred)
> dat1_2 <- dat1[dat1$pid%%20 == 0,]
> dat1_2$treat <- factor(dat1_2$treat, labels=c("Control", "Treatment"))
> ggplot(data=dat1_2, aes(x=time, y=stress, group=pid, colour=treat)) + geom_line()
```

การเปลี่ยนแปลงของคะแนนดิบคนไข้ทีละคน โดยแบ่งกลุ่มบำบัดและกลุ่มควบคุม



```
> datl <- split(dat1, dat1$pid)
> stresspred <- lapply(datl, function(x) { predict(lm(stress ~ timebase + timetreat + timefollow, data=x)) })
> dat1$stress2 <- do.call(c, stresspred)
> dat1_2 <- dat1[dat1$pid%%20 == 0,]
> dat1_2$treat <- factor(dat1_2$treat, labels=c("Control", "Treatment"))
> ggplot(data=dat1_2, aes(x=time, y=stress2, group=pid, colour=treat)) + geom_line()
```

การเปลี่ยนแปลงโดยดูจากค่าความเครียดที่ทำนายได้ จากโมเดลการเปลี่ยนแปลงเชิงเส้นช่วงย่อย ของคนไข้ทีละคนโดยแบ่งกลุ่มบำบัดและกลุ่มควบคุม



ตรวจสอบว่าการเปลี่ยนแปลงในแต่ละช่วงนั้น แตกต่างกันระหว่างบุคคลหรือไม่

อันดับแรก ตรวจสอบว่าการเปลี่ยนแปลงเชิงเส้นช่วงเก็บฐาน แตกต่างกันระหว่างกลุ่ม หรือไม่ พบว่าเมื่อกำหนดให้อิทธิพลของตัวแปรเป็นค่าสุ่มแล้ว ไม่สามารถประมาณค่าได้

```
> out1d <- lmer(stress ~ 1 + timebase + timetreat
+                 + timefollow + (1 + timebase|pid),
+                 data=dat1, REML=FALSE,
+                 control = lmerControl(optimizer ="Nelder_Mead"))
boundary (singular) fit: see ?isSingular
```

เนื่องจากผลการเปลี่ยนแปลงเชิงเส้นช่วงเก็บฐาน ไม่ถึงระดับนัยสำคัญ และไม่ใช่สิ่งที่ผู้วิจัยสนใจตั้งแต่ต้น จึงนำตัวแปรนี้ออกจากโมเดลในอนาคต

ตรวจสอบว่าการเปลี่ยนแปลงในแต่ละช่วงนั้น แตกต่างกันระหว่างบุคคลหรือไม่

```
> out1e <- lmer(stress ~ 1 + timebase + timetreat
+               + timefollow + (1 + timetreat|pid),
+               data=dat1, REML=FALSE,
+               control = lmerControl(optimizer ="Nelder_Mead"))
> anova(out1c, out1e)
Data: dat1
Models:
out1c: stress ~ 1 + timebase + timetreat + timefollow + (1 | pid)
out1e: stress ~ 1 + timebase + timetreat + timefollow + (1 + timetreat |
out1e:     pid)
      Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
out1c  6 102693 102738 -51341   102681
out1e  8 101783 101843 -50884   101767 913.99      2  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

การเปลี่ยนแปลงเชิงเส้นช่วงทดลอง แตกต่างกันระหว่างกลุ่มอย่างมีนัยสำคัญ

ตรวจสอบว่าการเปลี่ยนแปลงในแต่ละช่วงนั้น แตกต่างกันระหว่างบุคคลหรือไม่

```
> out1f <- lmer(stress ~ 1 + timebase + timetreat
+               + timefollow + (1 + timefollow|pid),
+               data=dat1, REML=FALSE,
+               control = lmerControl(optimizer ="Nelder_Mead"))
> anova(out1c, out1f)
Data: dat1
Models:
out1c: stress ~ 1 + timebase + timetreat + timefollow + (1 | pid)
out1f: stress ~ 1 + timebase + timetreat + timefollow + (1 + timefollow |
out1f:     pid)
      Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
out1c  6 102693 102738 -51341   102681
out1f  8 102450 102510 -51217   102434 247.44      2  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

การเปลี่ยนแปลงเชิงเส้นช่วงติดตามผล แตกต่างกันระหว่างกลุ่มอย่างมีนัยสำคัญ

ด้วยเหตุนี้ จึงสร้างโมเดลใหม่ ให้ช่วงทดลองและช่วงติดตามผลมีอิทธิพลสุ่ม

```
> out1g <- lmer(stress ~ 1 + timetreat + timefollow
+               + (1 + timetreat+ timefollow|pid),
+               data=dat1, REML=FALSE,
+               control = lmerControl(optimizer ="Nelder_Mead"))
> summary(out1g)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: stress ~ 1 + timetreat + timefollow + (1 + timetreat + t
   Data: dat1
Control: lmerControl(optimizer = "Nelder_Mead")

     AIC      BIC   logLik deviance df.resid
101766.7 101842.1 -50873.3 101746.7    13990

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.0056 -0.6302  0.0125  0.6550  3.2871

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pid      (Intercept)  0.3362  0.5799
          timetreat    1.3151  1.1468    0.20
          timefollow   1.8105  1.3455   -0.67 -0.31
 Residual             74.5464  8.6340
Number of obs: 14000, groups:  pid, 1000

Fixed effects:
            Estimate Std. Error t value
(Intercept) 52.40588    0.11110 471.680
timetreat   -0.88024    0.05103 -17.248
timefollow   0.58704    0.10957   5.358
```

$$\sigma^2 = 74.54$$

$$\tau_{00} = 0.34 \qquad \rho_{01} = .20$$

$$\tau_{11} = 1.32 \qquad \rho_{02} = -.67$$

$$\tau_{22} = 1.81 \qquad \rho_{12} = -.31$$

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + e_{ij}$$

$$\pi_{0j} = 52.41 + u_{0j}$$

$$\pi_{1j} = -0.88 + u_{1j}$$

$$\pi_{2j} = 0.59 + u_{2j}$$

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + e_{ij} \qquad \sigma^2 = 74.54$$

$$\pi_{0j} = 52.41 + u_{0j} \qquad \tau_{00} = 0.34 \qquad \rho_{01} = .20$$

$$\pi_{1j} = -0.88 + u_{1j} \qquad \tau_{11} = 1.32 \qquad \rho_{02} = -.67$$

$$\pi_{2j} = 0.59 + u_{2j} \qquad \tau_{22} = 1.81 \qquad \rho_{12} = -.31$$

ความเครียดช่วงก่อนบำบัด (เฉลี่ยข้อมูลการวัดครั้งที่ 1-5) เฉลี่ยคนไข้ทุกคนเท่ากับ 52.41 คะแนน มีช่วงเชื่อมั่น 95% เท่ากับ $52.41 \pm 1.96 \times \sqrt{0.34} = (51.27, 53.54)$ คะแนน

ในช่วงบำบัด คนไข้ทุกคนมีค่าเฉลี่ยความเครียดน้อยลง วันละ 0.88 คะแนน มีช่วงเชื่อมั่น 95% เท่ากับ $-0.88 \pm 1.96 \times \sqrt{1.32} = (-3.13, 1.37)$ คะแนนต่อวัน

ในช่วงติดตามผล คนไข้ทุกคนมีค่าเฉลี่ยความเรียดมากขึ้น วันละ 0.59 คะแนน มีช่วงเชื่อมั่น 95% เท่ากับ $0.59 \pm 1.96 \times \sqrt{1.81} = (-2.96, 4.14)$ คะแนนต่อวัน

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + e_{ij}$$

$$\pi_{0j} = 52.41 + u_{0j}$$

$$\pi_{1j} = -0.88 + u_{1j}$$

$$\pi_{2j} = 0.59 + u_{2j}$$

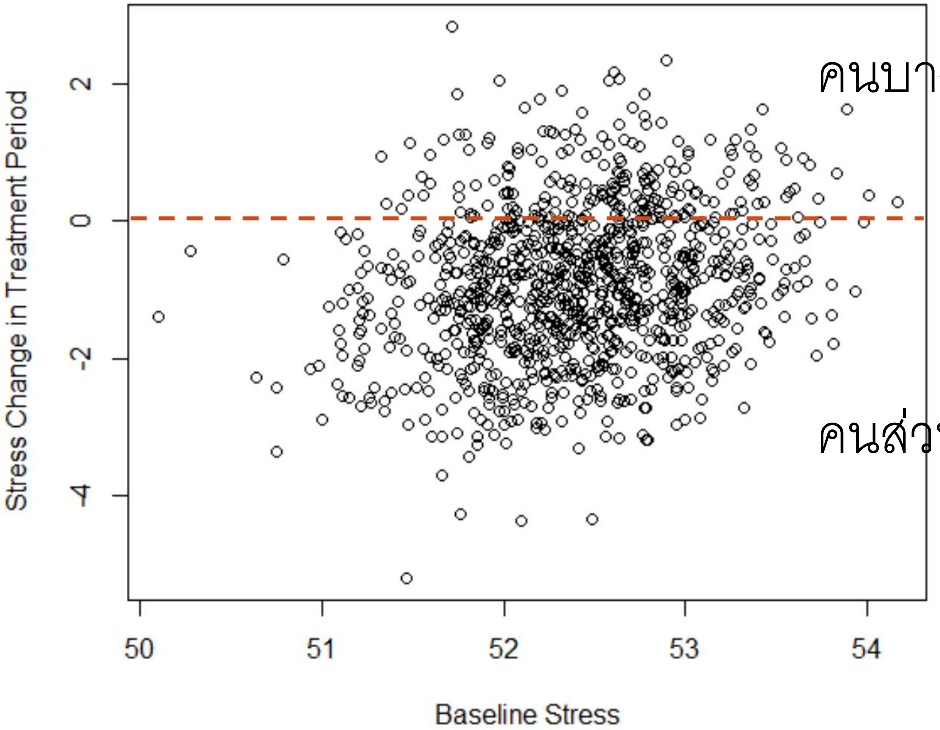
$$\sigma^2 = 74.54$$

$$\tau_{00} = 0.34 \qquad \rho_{01} = .20$$

$$\tau_{11} = 1.32 \qquad \rho_{02} = -.67$$

$$\tau_{22} = 1.81 \qquad \rho_{12} = -.31$$

หากความเครียดช่วงก่อนบำบัดมีค่ามาก จะมีแนวโน้มที่ความลดลงของความเครียดมีค่าเบาลงในช่วงการบำบัด โดยมีค่าสหสัมพันธ์เท่ากับ .2



คนบางส่วนมีความเครียดมากขึ้น

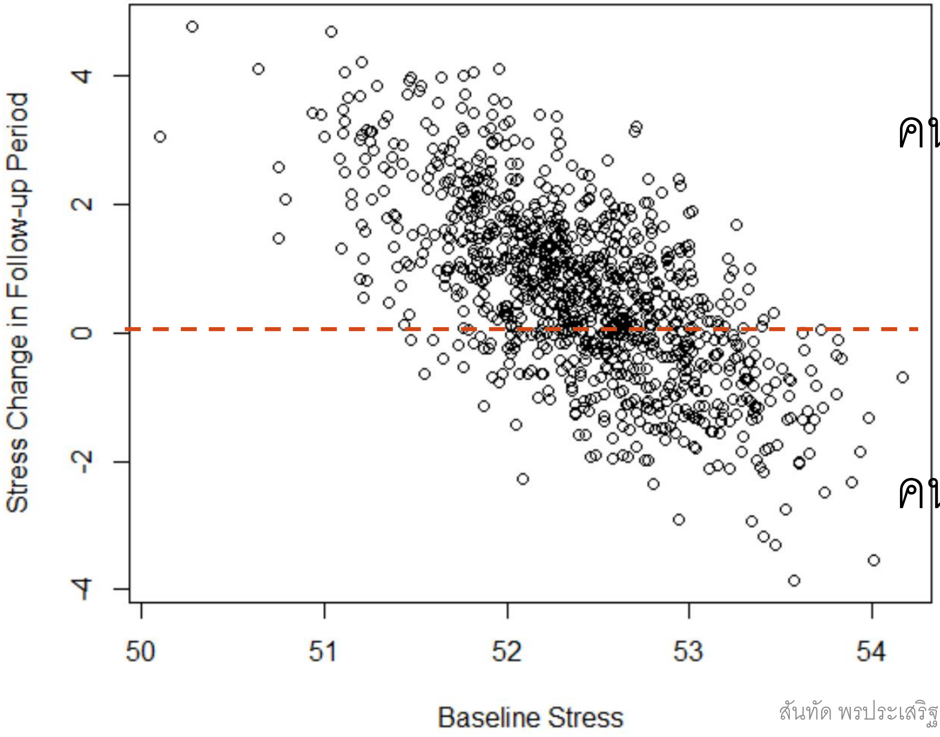
คนที่เครียดมากขึ้นจะมีแนวโน้มที่ความเครียดเริ่มต้นสูงกว่า

คนส่วนใหญ่ความเครียดน้อยลงช่วงบำบัด

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + e_{ij}$$
$$\pi_{0j} = 52.41 + u_{0j}$$
$$\pi_{1j} = -0.88 + u_{1j}$$
$$\pi_{2j} = 0.59 + u_{2j}$$

$$\sigma^2 = 74.54$$
$$\tau_{00} = 0.34 \qquad \rho_{01} = .20$$
$$\tau_{11} = 1.32 \qquad \rho_{02} = -.67$$
$$\tau_{22} = 1.81 \qquad \rho_{12} = -.31$$

หากความเครียดช่วงก่อนบำบัดมีค่ามาก จะมีแนวโน้มที่ช่วงติดตามจะมีความเครียดลดลง แต่หากความเครียดก่อนบำบัดมีค่าน้อย จะมีแนวโน้มความเครียดช่วงติดตามจะมีความเครียดเพิ่มขึ้น ค่าสหสัมพันธ์เท่ากับ -.67



คนส่วนใหญ่ความเครียดสูงขึ้น

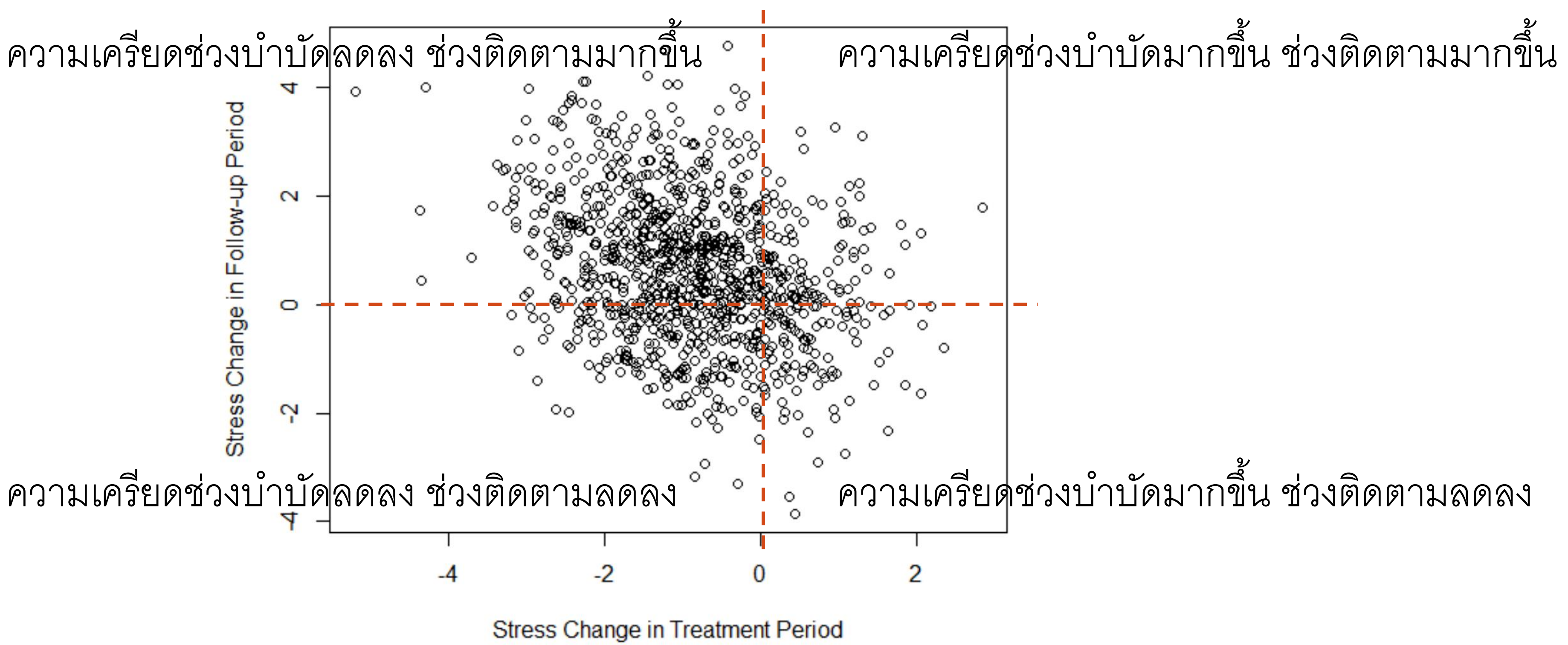
คนที่เครียดมากขึ้นจะมีแนวโน้มที่ความเครียดเริ่มต้นต่ำกว่า

คนบางส่วนมีความเครียดน้อยลง

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + e_{ij}$$
$$\pi_{0j} = 52.41 + u_{0j}$$
$$\pi_{1j} = -0.88 + u_{1j}$$
$$\pi_{2j} = 0.59 + u_{2j}$$

$$\sigma^2 = 74.54$$
$$\tau_{00} = 0.34 \qquad \rho_{01} = .20$$
$$\tau_{11} = 1.32 \qquad \rho_{02} = -.67$$
$$\tau_{22} = 1.81 \qquad \rho_{12} = -.31$$

หากความเครียดช่วงบำบัดลดลง จะมีแนวโน้มที่ความเครียดช่วงติดตามผลมากขึ้น ค่าสหสัมพันธ์อยู่ที่ -.31



แปลงช่วงนอนการนอนออกเป็น 2 ระดับ คือ ระดับวันและระดับบุคคล

เปลี่ยนหน่วยจากนาทีเป็นชั่วโมง

```
> dat1$sleephrs <- dat1$sleep / 60
> dat1$avesleep <- ave(dat1$sleephrs, dat1$pid)
> dat1$diffsleep <- dat1$sleephrs - dat1$avesleep
```

หาค่าเฉลี่ยของตัวแปรระดับบุคคล เพื่อนำไปย้ายศูนย์กลาง

```
> dat1a <- dat1[!duplicated(dat1$pid),]
> round(apply(dat1a, 2, mean), 3)
    rowid       pid      time     stress     treat     sleep
 6994.000   500.500     1.000     52.543     0.500   424.507
      age     timec  timebase  timetreat timefollow  sleephrs
   36.638     0.000     0.000      0.000     0.000     7.075
 avesleep diffsleep
    7.060     0.015
```

หากย้ายศูนย์กลางของตัวแปรดัมมี่เท่ากับค่าเฉลี่ยแล้ว

จะเรียกตัวแปรดัมมี่ว่ามีการใส่รหัสแบบอิทธิพล (Effect Coding)

```
> out1h <- lmer(stress ~ 1 + timetreat + timefollow + diffsleep
+               + I(avesleep - 7.075) + I(treat - 0.5) + I(age - 36.638)
+               + (1 + timetreat+ timefollow|pid), data=dat1, REML=FALSE,
+               control = lmerControl(optimizer ="Nelder_Mead"))
> summary(out1h)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: stress ~ 1 + timetreat + timefollow + diffsleep + I(avesleep -
    7.075) + I(treat - 0.5) + I(age - 36.638) + (1 + timetreat +      time
   Data: dat1
Control: lmerControl(optimizer = "Nelder_Mead")

     AIC      BIC   logLik deviance df.resid
101443.5 101549.2 -50707.7 101415.5    13986

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.0920 -0.6324  0.0072  0.6530  3.2911

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pid      (Intercept)  2.446   1.564
          timetreat    1.306   1.143    -0.65
          timefollow   1.713   1.309     0.16 -0.30
 Residual             73.511   8.574
Number of obs: 14000, groups:  pid, 1000

Fixed effects:
                    Estimate Std. Error t value
(Intercept)         52.41799    0.11955 438.465
timetreat           -0.88473    0.05077 -17.427
timefollow           0.59457    0.10847   5.481
diffsleep           -1.91626    0.14072 -13.617
I(avesleep - 7.075)  0.18586    0.12980   1.432
I(treat - 0.5)      -2.82121    0.18266 -15.445
I(age - 36.638)      0.01326    0.01724   0.769
```

$\sigma^2 = 73.51$

$\tau_{00} = 2.45$ $\qquad \rho_{01} = -.65$

$\tau_{11} = 1.31$ $\qquad \rho_{02} = .16$

$\tau_{22} = 1.71$ $\qquad \rho_{12} = -.30$

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + \pi_{3j}X_{ij} + e_{ij}$$

$$\pi_{0j} = 52.42 - 2.82(W_{1j} - 0.5) + 0.013(W_{2j} - 36.638) + 0.186(\bar{X}_{2ij} - 7.075) + u_{0j}$$

$$\pi_{1j} = -0.88 + u_{1j} \qquad \pi_{2j} = 0.59 + u_{2j} \qquad \pi_{3j} = -1.92$$

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + \pi_{3j}X_{ij} + e_{ij}$$

$$\pi_{0j} = 52.42 - 2.82(W_{1j} - 0.5) + 0.013(W_{2j} - 36.638) + 0.186(\bar{X}_{2ij} - 7.075) + u_{0j}$$

$$\pi_{1j} = -0.88 + u_{1j} \qquad \pi_{2j} = 0.59 + u_{2j} \qquad \pi_{3j} = -1.92$$

ความเครียดช่วงก่อนบำบัด (เฉลี่ยข้อมูลการวัดครั้งที่ 1-5) เฉลี่ยคนไข้ทุกคนเท่ากับ 52.42 คะแนน เมื่อตัวแปรการเข้าบำบัด อายุของผู้เข้าร่วม และจำนวนชั่วโมงการนอน เท่ากับค่าเฉลี่ย

ในช่วงบำบัด คนไข้ทุกคนมีค่าเฉลี่ยความเครียดน้อยลง วันละ 0.88 คะแนน

ในช่วงติดตามผล คนไข้ทุกคนมีค่าเฉลี่ยความเครียดมากขึ้น วันละ 0.59 คะแนน

ภายในคนไข้คนเดียวกัน หากคนดังกล่าวมีชั่วโมงการนอนน้อยลง 1 ชั่วโมง จะมีความเครียดน้อยลง 1.92 แต้ม

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + \pi_{3j}X_{ij} + e_{ij}$$

$$\pi_{0j} = 52.42 - 2.82(W_{1j} - 0.5) + 0.013(W_{2j} - 36.638) + 0.186(\bar{X}_{2ij} - 7.075) + u_{0j}$$

$$\pi_{1j} = -0.88 + u_{1j} \qquad \pi_{2j} = 0.59 + u_{2j} \qquad \pi_{3j} = -1.92$$

เปรียบเทียบระหว่างกลุ่มบำบัดและกลุ่มควบคุม กลุ่มบำบัดมีความเครียดช่วงเก็บฐานน้อยกว่ากลุ่มควบคุม 2.82 คะแนน เมื่อควบคุมตัวแปรอายุ และชั่วโมงการนอนเฉลี่ยให้คงที่

หากอายุมากขึ้น 1 ปี ความเครียดช่วงเก็บฐานเพิ่มขึ้น 0.013 คะแนน เมื่อควบคุมเงื่อนไขกลุ่มทดลอง และชั่วโมงการนอนให้คงที่ แต่ผลนี้ไม่ถึงระดับนัยสำคัญ

เปรียบเทียบระหว่างคนไข้สองคน ที่มีชั่วโมงการนอนเฉลี่ยแตกต่างกัน 1 ชั่วโมง คนที่ชั่วโมงการนอนสูงกว่าจะมีความเครียดมากกว่า 0.186 คะแนน เมื่อควบคุมเงื่อนไขกลุ่มทดลอง และอายุให้คงที่ แต่ผลนี้ไม่ถึงระดับนัยสำคัญ

อีกคำสั่งหนึ่งที่ช่วยให้การวิเคราะห์ด้วย `lmer` หาคำตอบ คือ

```
control = lmerControl(calc.derivs = FALSE))
```

cran.r-project.org/web/packages/lme4/vignettes/lmerperf.html

```
library("lme4")
```

lme4 Performance Tips

- use control = `[g]lmerControl(calc.derivs = FALSE)` to turn off the time consuming derivative calculation that is performed after the optmization is finished, e.g.

```
lmer(y ~ service * dept + (1|s) + (1|d), InstEval,
     control = lmerControl(calc.derivs = FALSE))
```

Note that this will disable some of the convergence tests, as well as (for `glmer` only) making `lme4` use a less accurate approximation to compute the standard errors of the fixed effects.

ในกรณีในคาบเรียนนี้ ที่ใช้เฉพาะตัวแปรตามแบบต่อเนื่อง การคำนวณ *SE* ยังไม่มีผล

```
> out1i <- lmer(stress ~ 1 + timetreat + timefollow + diffsleep
+               + timetreat*I(avesleep - 7.075) + timetreat*I(treat - 0.5)
+               + timetreat*I(age - 36.638) + timefollow*I(avesleep - 7.075)
+               + timefollow*I(treat - 0.5) + timefollow*I(age - 36.638)
+               + (1 + timetreat+ timefollow|pid), data=dat1, REML=FALSE,
+               control = lmerControl(optimizer ="Nelder_Mead", calc.derivs = FALSE))
> summary(out1i)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: stress ~ 1 + timetreat + timefollow + diffsleep + timetreat *
    I(avesleep - 7.075) + timetreat * I(treat - 0.5) + timetreat *
    I(age - 36.638) + timefollow * I(avesleep - 7.075) + timefollow *
    I(treat - 0.5) + timefollow * I(age - 36.638) + (1 + timetreat +      timefollow | pid)
   Data: dat1
Control: lmerControl(optimizer = "Nelder_Mead", calc.derivs = FALSE)

      AIC       BIC    logLik  deviance df.resid
 101023.0  101173.9  -50491.5  100983.0    13980

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.0674 -0.6351  0.0053  0.6621  3.3091

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pid      (Intercept)  0.3553  0.5960
          timetreat    0.4344  0.6591    0.46
          timefollow   1.3450  1.1598   -0.79  0.14
 Residual             73.5184  8.5743
Number of obs: 14000, groups:  pid, 1000
```

$\sigma^2 = 73.52$

$\tau_{00} = 0.36$ $\qquad \rho_{01} = .46$

$\tau_{11} = 0.43$ $\qquad \rho_{02} = -.79$

$\tau_{22} = 1.35$ $\qquad \rho_{12} = .14$

```
Fixed effects:
                             Estimate Std. Error t value
(Intercept)                 52.418891   0.110471 474.503
timetreat                   -0.885219   0.041316 -21.426
timefollow                   0.593959   0.106793   5.562
diffsleep                   -1.892950   0.139462 -13.573
I(avesleep - 7.075)          0.251976   0.156970   1.605
I(treat - 0.5)               0.059906   0.220894   0.271
I(age - 36.638)              0.013452   0.020844   0.645
timetreat:I(avesleep - 7.075) -0.035437 0.058707  -0.604
timetreat:I(treat - 0.5)    -1.864007   0.082616 -22.562
timetreat:I(age - 36.638)    0.002091   0.007796   0.268
timefollow:I(avesleep - 7.075) -0.033808 0.151753 -0.223
timefollow:I(treat - 0.5)    1.188626   0.213551   5.566
timefollow:I(age - 36.638)  -0.018400   0.020154  -0.913
```

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + \pi_{3j}X_{ij} + e_{ij}$$

$$\pi_{0j} = 52.42 + 0.060(W_{1j} - 0.5) + 0.013(W_{2j} - 36.638) + 0.252(\bar{X}_{2ij} - 7.075) + u_{0j}$$

$$\pi_{1j} = -0.88 - 1.864(W_{1j} - 0.5) + 0.002(W_{2j} - 36.638) - 0.035(\bar{X}_{2ij} - 7.075) + u_{1j}$$

$$\pi_{2j} = 0.59 + 1.189(W_{1j} - 0.5) - 0.018(W_{2j} - 36.638) - 0.034(\bar{X}_{2ij} - 7.075) + u_{2j}$$

$$\pi_{3j} = -1.89$$

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + \pi_{3j}X_{ij} + e_{ij}$$

$$\pi_{0j} = 52.42 + 0.060(W_{1j} - 0.5) + 0.013(W_{2j} - 36.638) + 0.252(\bar{X}_{2ij} - 7.075) + u_{0j}$$

$$\pi_{1j} = -0.88 - 1.864(W_{1j} - 0.5) + 0.002(W_{2j} - 36.638) - 0.035(\bar{X}_{2ij} - 7.075) + u_{1j}$$

$$\pi_{2j} = 0.59 + 1.189(W_{1j} - 0.5) - 0.018(W_{2j} - 36.638) - 0.034(\bar{X}_{2ij} - 7.075) + u_{2j}$$

$$\pi_{3j} = -1.89$$

ความเครียดช่วงก่อนบำบัด (เฉลี่ยข้อมูลการวัดครั้งที่ 1-5) เฉลี่ยคนไข้ทุกคนเท่ากับ 52.42 คะแนน เมื่อตัวแปรการเข้าบำบัด อายุของผู้เข้าร่วม และจำนวนชั่วโมงการนอน เท่ากับค่าเฉลี่ย

เปรียบเทียบระหว่างกลุ่มบำบัดและกลุ่มควบคุม กลุ่มบำบัดมีความเครียดช่วงเก็บฐานน้อยกว่ากลุ่มควบคุม 0.06 คะแนน เมื่อตัวแปรอายุ และชั่วโมงการนอนเฉลี่ยเท่ากับค่าเฉลี่ยรวมแต่ผลไม่ถึงระดับนัยสำคัญ

หากอายุมากขึ้น 1 ปี ความเครียดช่วงเก็บฐานเพิ่มขึ้น 0.013 คะแนน เมื่อค่าเงื่อนไขกลุ่มทดลอง และชั่วโมงการนอนให้เท่ากับค่าเฉลี่ย แต่ผลนี้ไม่ถึงระดับนัยสำคัญ

เปรียบเทียบระหว่างคนไข้สองคน ที่มีชั่วโมงการนอนเฉลี่ยแตกต่างกัน 1 ชั่วโมง คนที่ชั่วโมงการนอนสูงกว่า จะมีความเครียดมากกว่า 0.252 คะแนน เมื่อค่าเงื่อนไขกลุ่มทดลอง และอายุเท่ากับค่าเฉลี่ย แต่ผลนี้ไม่ถึงระดับนัยสำคัญ

ภายในคนไข้ที่เงื่อนไขกลุ่มทดลอง อายุ และจำนวนการนอนเฉลี่ยเท่ากับค่าเฉลี่ย หากคนดังกล่าวมีชั่วโมงการนอนน้อยลง 1 ชั่วโมง จะมีความเครียดน้อยลง 1.89 แต้ม

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + \pi_{3j}X_{ij} + e_{ij}$$

$$\pi_{0j} = 52.42 + 0.060(W_{1j} - 0.5) + 0.013(W_{2j} - 36.638) + 0.252(\bar{X}_{2ij} - 7.075) + u_{0j}$$

$$\pi_{1j} = -0.88 - 1.864(W_{1j} - 0.5) + 0.002(W_{2j} - 36.638) - 0.035(\bar{X}_{2ij} - 7.075) + u_{1j}$$

$$\pi_{2j} = 0.59 + 1.189(W_{1j} - 0.5) - 0.018(W_{2j} - 36.638) - 0.034(\bar{X}_{2ij} - 7.075) + u_{2j}$$

$$\pi_{3j} = -1.89$$

ในช่วงบำบัด คนไข้ทุกคนมีค่าเฉลี่ยความเครียดน้อยลง วันละ 0.88 คะแนน เมื่อตัวแปรการเข้าบำบัด อายุของผู้เข้าร่วม และจำนวนชั่วโมงการนอน เท่ากับค่าเฉลี่ย

เปรียบเทียบระหว่างกลุ่มบำบัดและกลุ่มควบคุม กลุ่มบำบัดมีความเครียดลดลงในช่วงบำบัด ในอัตราที่เร็วกว่ากลุ่มควบคุมอยู่ 1.86 คะแนนต่อวัน เมื่อควบคุมตัวแปรอายุ และชั่วโมงการนอนเฉลี่ยให้คงที่

หากอายุมากขึ้น 1 ปี อัตราการลดลงของความเครียดในช่วงบำบัด มีอัตราที่น้อยลง 0.002 คะแนนต่อวัน เมื่อควบคุมเงื่อนไขกลุ่มทดลอง และชั่วโมงการนอนให้คงที่ แต่ผลนี้ไม่ถึงระดับนัยสำคัญ

เปรียบเทียบระหว่างคนไข้สองคน ที่มีชั่วโมงการนอนเฉลี่ยแตกต่างกัน 1 ชั่วโมง คนที่ชั่วโมงการนอนสูงกว่า จะมีอัตราการลดลงของความเครียดที่เร็วขึ้นในช่วงบำบัด 0.035 คะแนนต่อวัน เมื่อควบคุมเงื่อนไขกลุ่มทดลอง และอายุให้คงที่ แต่ผลนี้ไม่ถึงระดับนัยสำคัญ

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{Tij} + \pi_{2j}T_{Fij} + \pi_{3j}X_{ij} + e_{ij}$$

$$\pi_{0j} = 52.42 + 0.060(W_{1j} - 0.5) + 0.013(W_{2j} - 36.638) + 0.252(\bar{X}_{2ij} - 7.075) + u_{0j}$$

$$\pi_{1j} = -0.88 - 1.864(W_{1j} - 0.5) + 0.002(W_{2j} - 36.638) - 0.035(\bar{X}_{2ij} - 7.075) + u_{1j}$$

$$\pi_{2j} = 0.59 + 1.189(W_{1j} - 0.5) - 0.018(W_{2j} - 36.638) - 0.034(\bar{X}_{2ij} - 7.075) + u_{2j}$$

$$\pi_{3j} = -1.89$$

ในช่วงติดตามผล คนไข้ทุกคนมีค่าเฉลี่ยความเครียดมากขึ้น วันละ 0.59 คะแนน
เมื่อตัวแปรการเข้าบำบัด อายุของผู้เข้าร่วม และจำนวนชั่วโมงการนอน เท่ากับค่าเฉลี่ย

เปรียบเทียบระหว่างกลุ่มบำบัดและกลุ่มควบคุม กลุ่มบำบัดมีความเครียดเพิ่มขึ้นในช่วงติดตามผล ในอัตราที่มากกว่ากว่ากลุ่มควบคุมอยู่ 1.19 คะแนนต่อวัน เมื่อควบคุมตัวแปรอายุ และชั่วโมงการนอนเฉลี่ยให้คงที่

หากอายุมากขึ้น 1 ปี อัตราการเพิ่มขึ้นของความเครียดในช่วงติดตาม มีอัตราที่น้อยลง 0.018 คะแนนต่อวัน เมื่อควบคุมเงื่อนไขกลุ่มทดลอง และชั่วโมงการนอนให้คงที่ แต่ผลนี้ไม่ถึงระดับนัยสำคัญ

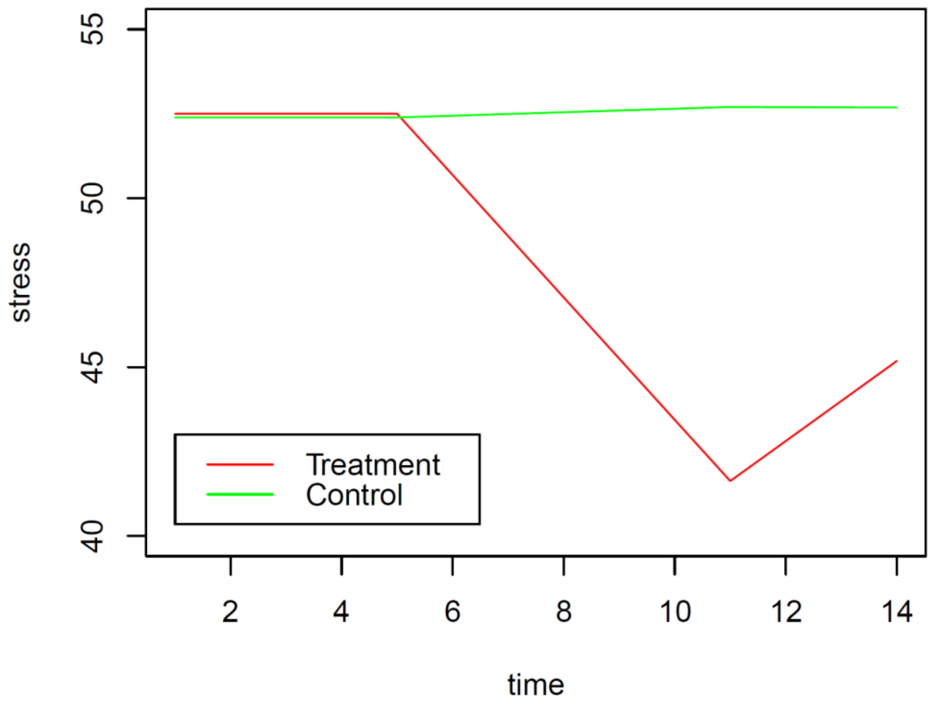
เปรียบเทียบระหว่างคนไข้สองคน ที่มีชั่วโมงการนอนเฉลี่ยแตกต่างกัน 1 ชั่วโมง คนที่ชั่วโมงการนอนสูงกว่า จะมีอัตราการเพิ่มขึ้นของความเครียดในช่วงติดตามช้าลง 0.034 คะแนนต่อวัน เมื่อควบคุมเงื่อนไขกลุ่มทดลอง และอายุให้คงที่ แต่ผลนี้ไม่ถึงระดับนัยสำคัญ

แทนตัวแปรอื่นนอกจากเงื่อนไขการบำบัดให้เท่ากับค่าเฉลี่ย เพื่อดูการเปลี่ยนแปลงของความเครียด ในแต่ละเงื่อนไขการบำบัด

$$Y_{ij} = 52.42 + 0.060(W_{1j} - 0.5) - 0.88T_{Tij} - 1.864(W_{1j} - 0.5)T_{Tij} + 0.59T_{Fij} + 1.189(W_{1j} - 0.5)T_{Fij}$$

ให้ $W_1 = 1$ $\qquad Y_{ij} = 52.499 - 1.812T_{Tij} + 1.185T_{Fij}$

ให้ $W_1 = 0$ $\qquad Y_{ij} = 52.389 + 0.052T_{Tij} - 0.005T_{Fij}$



```
> treatcval <- c(0, 1) - 0.5
> library(interactions)
> ss11 <- sim_slopes(model=out1in, pred=timetreat, modx=treatc, modx.values=treatcval)

> ss11
JOHNSON-NEYMAN INTERVAL

When treatc is OUTSIDE the interval [-0.54, -0.42], the slope of timetreat
is p < .05.

Note: The range of observed values of treatc is [-0.50, 0.50]

SIMPLE SLOPES ANALYSIS

Slope of timetreat when treatc = -0.50:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.05   0.06     0.81   0.42

Slope of timetreat when treatc =  0.50:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -1.82   0.06   -31.06   0.00
```

```
> ss12 <- sim_slopes(model=out1in, pred=timefollow, modx=treatc, modx.values=treatcval)
```

```
> ss12
JOHNSON-NEYMAN INTERVAL

When treatc is OUTSIDE the interval [-0.85, -0.30], the slope of
timefollow is p < .05.

Note: The range of observed values of treatc is [-0.50, 0.50]

SIMPLE SLOPES ANALYSIS

Slope of timefollow when treatc = -0.50:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  0.00   0.15     0.00   1.00

Slope of timefollow when treatc =  0.50:

  Est.   S.E.   t val.      p
------ ------ -------- ------
  1.19   0.15     7.85   0.00
```

```
> timetreatval <- c(0, 1, 2, 3, 4, 5, 6)
> timefollowval <- c(0, 1, 2, 3)
> ss13 <- sim_slopes(model=out1in, pred=treatc, modx=timetreat, mod2=timefollow,
modx.values=timetreatval, mod2.values=timefollowval)
```

```
> ss13
███████████████████ While timefollow (2nd moderator) = 3.00 ██████████████████

JOHNSON-NEYMAN INTERVAL

When timetreat is OUTSIDE the interval [1.27, 2.57], the slope of
treatc is p < .05.

Note: The range of observed values of timetreat is [0.00, 6.00]

SIMPLE SLOPES ANALYSIS

Slope of treatc when timetreat = 6.00:

   Est.   S.E.   t val.      p
------- ------ -------- ------
  -7.56   0.57   -13.32   0.00
```

███████████████████ While timefollow (2nd moderator) = 2.00 ██████████████████

JOHNSON-NEYMAN INTERVAL

When timetreat is OUTSIDE the interval [0.82, 1.75], the slope of treatc is p < .05.

Note: The range of observed values of timetreat is [0.00, 6.00]

SIMPLE SLOPES ANALYSIS

Slope of treatc when timetreat = 6.00:

Est.	S.E.	t val.	p
-8.75	0.42	-20.60	0.00

███████████████████ While timefollow (2nd moderator) = 1.00 ███████████████████

JOHNSON-NEYMAN INTERVAL

When timetreat is OUTSIDE the interval [0.35, 0.96], the slope of treatc is $p < .05$.

Note: The range of observed values of timetreat is [0.00, 6.00]

SIMPLE SLOPES ANALYSIS

Slope of treatc when timetreat = 6.00:

Est.	S.E.	t val.	p
-9.94	0.37	-27.11	0.00

███████████████████ While timefollow (2nd moderator) = 0.00 ██████████████████

JOHNSON-NEYMAN INTERVAL

When timetreat is OUTSIDE the interval [-0.21, 0.25], the slope of treatc is p < .05.

Note: The range of observed values of timetreat is [0.00, 6.00]

SIMPLE SLOPES ANALYSIS

Slope of treatc when timetreat = 6.00:

Est.	S.E.	t val.	p
-11.12	0.42	-26.37	0.00

Slope of treatc when timetreat = 5.00:

Est.	S.E.	t val.	p
-9.26	0.35	-26.49	0.00

Slope of treatc when timetreat = 4.00:

Est.	S.E.	t val.	p
-7.40	0.28	-26.13	0.00

Slope of treatc when timetreat = 3.00:

Est.	S.E.	t val.	p
-5.53	0.23	-24.31	0.00

Slope of treatc when timetreat = 2.00:

Est.	S.E.	t val.	p
-3.67	0.16	-22.45	0.00

Slope of treatc when timetreat = 1.00:

Est.	S.E.	t val.	p
-1.80	0.19	-9.49	0.00

Slope of treatc when timetreat = 0.00:

Est.	S.E.	t val.	p
0.06	0.22	0.27	0.78

จากสมการ

$$Y_{ij} = 52.42 + 0.060(W_{1j} - 0.5) - 0.88T_{Tij} - 1.864(W_{1j} - 0.5)T_{Tij}$$
$$+0.59T_{Fij} + 1.189(W_{1j} - 0.5)T_{Fij}$$

สามารถตรวจสอบปฏิสัมพันธ์ได้ 2 แบบ คือ

1. ตรวจสอบว่าความชันของ T_T และ T_F ว่าแตกต่างจาก 0 อย่างมีนัยสำคัญหรือไม่ ในกลุ่มบำบัด $(W_1 = 1)$ และกลุ่มควบคุม $(W_1 = 0)$

$$Y_{ij} = 52.42 + 0.060(W_{1j} - 0.5) + \left(-0.88T_{Tij} - 1.864(W_{1j} - 0.5)\right)T_{Tij}$$
$$+\left(0.59T_{Fij} + 1.189(W_{1j} - 0.5)\right)T_{Fij}$$

W_1	$W_1 - 0.5$	จุดตัดอย่างง่าย	ความชันอย่างง่ายของ T_T	ความชันอย่างง่ายของ T_F
1	0.5	52.52 + (0.060 x (1 – 0.5)) = 52.50	-0.88 + (-1.86 x (1 – 0.5)) = -1.81	0.59 + (1.19 x (1 – 0.5)) = 1.19
0	-0.5	52.52 + (0.060 x (0 – 0.5)) = 52.39	-0.88 + (-1.86 x (0 – 0.5)) = 0.05	0.59 + (1.19 x (0 – 0.5)) = -0.01

W_1	$W_1 - 0.5$	จุดตัดอย่างง่าย	ความชันอย่างง่ายของ T_T	ความชันอย่างง่ายของ T_F
1	0.5	52.52 + (0.060 x (1 – 0.5)) = 52.50	-0.88 + (-1.86 x (1 – 0.5)) = -1.81	0.59 + (1.19 x (1 – 0.5)) = 1.19
0	-0.5	52.52 + (0.060 x (0 – 0.5)) = 52.39	-0.88 + (-1.86 x (0 – 0.5)) = 0.05	0.59 + (1.19 x (0 – 0.5)) = -0.01

```
> sumout1i <- summary(out1i)
> coef(sumout1i)
```

	Estimate	Intercept	Slope T	Slope F
(Intercept)	52.418891213	1	0	0
timetreat	-0.885219293	0	1	0
timefollow	0.593958519	0	0	1
diffsleep	-1.892949777	0	0	0
I(avesleep - 7.075)	0.251976002	0	0	0
I(treat - 0.5)	0.059905775	-0.5, 0.5	0	0
I(age - 36.638)	0.013452280	0	0	0
timetreat:I(avesleep - 7.075)	-0.035437335	0	0	0
timetreat:I(treat - 0.5)	-1.864006760	0	-0.5, 0.5	0
timetreat:I(age - 36.638)	0.002091415	0	0	0
timefollow:I(avesleep - 7.075)	-0.033807955	0	0	0
timefollow:I(treat - 0.5)	1.188625511	0	0	-0.5, 0.5
timefollow:I(age - 36.638)	-0.018400424	0	0	0

```
> psi <- coef(sumout1i)[,1]
> vpsi <- vcov(sumout1i)
> cvec1 <- c(1, 0, 0, 0, 0, 0.5, 0, 0, 0, 0, 0, 0, 0)
> cvec2 <- c(1, 0, 0, 0, 0, -0.5, 0, 0, 0, 0, 0, 0, 0)
> cvec3 <- c(0, 1, 0, 0, 0, 0, 0, 0, 0.5, 0, 0, 0, 0)
> cvec4 <- c(0, 1, 0, 0, 0, 0, 0, 0, -0.5, 0, 0, 0, 0)
> cvec5 <- c(0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0.5, 0)
> cvec6 <- c(0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, -0.5, 0)
> cmat <- cbind(cvec1, cvec2, cvec3, cvec4, cvec5, cvec6)
> est <- t(cmat) %*% psi
> se <- sqrt(diag(t(cmat) %*% vpsi %*% cmat))
> zval <- est/se
> pval <- round(pnorm(-abs(zval)) * 2, 6)
> cbind(est, se, zval, pval)
                            se
cvec1 52.4488441003 0.15621377 335.750457705 0.000000
cvec2 52.3889383256 0.15621178 335.371236171 0.000000
cvec3 -1.8172226734 0.05842283 -31.104665646 0.000000
cvec4  0.0467840870 0.05842488   0.800756198 0.423273
cvec5  1.1882712746 0.15101591   7.868517250 0.000000
cvec6 -0.0003542361 0.15101620  -0.002345683 0.998128
```

ในกลุ่มบำบัด มีความเครียดลดลง 1.82 แต้มต่อวัน ในช่วงบำบัด และมีความเครียดกลับมาสูงขึ้น 1.19 แต้มต่อวันในช่วงติดตาม

ในกลุ่มควบคุม มีความเครียดในช่วงบำบัดและช่วงติดตาม ไม่เปลี่ยนแปลงอย่างมีนัยสำคัญ

เงื่อนไขทดลอง	จุดตัดแบบง่าย			ความชันแบบง่ายช่วงบำบัด			ความชันแบบง่ายช่วงติดตาม		
	b	*SE*	*z*	*b*	*SE*	*z*	*b*	*SE*	*z*
กลุ่มบำบัด	52.50	0.16	335.75**	-1.82	0.06	-31.10**	1.19	0.15	7.87**
กลุ่มควบคุม	52.39	0.16	335.37**	0.05	0.06	0.80	0.00	0.15	-0.00

สามารถตรวจสอบปฏิสัมพันธ์ได้ 2 แบบ คือ

2. ในแต่ละช่วงเวลา ความเครียดแตกต่างระหว่างกลุ่มทดลองและกลุ่มควบคุมหรือไม่

$$Y_{ij} = 52.42 - 0.88T_{Tij} + 0.59T_{Fij} + (0.060 - 1.864T_{Tij} + 1.189T_{Fij})(W_{1j} - 0.5)$$

T	T_T	T_F	จุดตัดอย่างง่าย	ความชันอย่างง่าย
1-5	0	0	52.42 + (-0.88 x 0) + (0.59 x 0) = 52.42	0.060 + (-1.864 x 0) + (1.189 x 0) = 0.06
6	1	0	52.42 + (-0.88 x 1) + (0.59 x 0) = 51.54	0.060 + (-1.864 x 1) + (1.189 x 0) = -1.80
7	2	0	52.42 + (-0.88 x 2) + (0.59 x 0) = 50.66	0.060 + (-1.864 x 2) + (1.189 x 0) = -3.67
8	3	0	52.42 + (-0.88 x 3) + (0.59 x 0) = 49.78	0.060 + (-1.864 x 3) + (1.189 x 0) = -5.53
9	4	0	52.52 + (-0.88 x 4) + (0.59 x 0) = 48.90	0.060 + (-1.864 x 4) + (1.189 x 0) = -7.40
10	5	0	52.42 + (-0.88 x 5) + (0.59 x 0) = 48.02	0.060 + (-1.864 x 5) + (1.189 x 0) = -9.26
11	6	0	52.42 + (-0.88 x 6) + (0.59 x 0) = 47.14	0.060 + (-1.864 x 6) + (1.189 x 0) = -11.12
12	6	1	52.42 + (-0.88 x 6) + (0.59 x 1) = 47.73	0.060 + (-1.864 x 6) + (1.189 x 1) = -9.94
13	6	2	52.42 + (-0.88 x 6) + (0.59 x 2) = 48.32	0.060 + (-1.864 x 6) + (1.189 x 2) = -8.75
14	6	3	52.42 + (-0.88 x 6) + (0.59 x 3) = 48.91	0.060 + (-1.864 x 6) + (1.189 x 3) = -7.56

```
> sumout1i <- summary(out1i)
> coef(sumout1i)
```

	Estimate	Slope F
(Intercept)	52.418891213	0
timetreat	-0.885219293	0
timefollow	0.593958519	0
diffsleep	-1.892949777	0
I(avesleep - 7.075)	0.251976002	0
I(treat - 0.5)	0.059905775	1
I(age - 36.638)	0.013452280	0
timetreat:I(avesleep - 7.075)	-0.035437335	0
timetreat:I(treat - 0.5)	-1.864006760	0, 1, 2, 3, 4, 5, 6, 6, 6, 6
timetreat:I(age - 36.638)	0.002091415	0
timefollow:I(avesleep - 7.075)	-0.033807955	0
timefollow:I(treat - 0.5)	1.188625511	0, 0, 0, 0, 0, 0, 0, 1, 2, 3
timefollow:I(age - 36.638)	-0.018400424	0

```
> cvec1 <- c(0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0)
> cvec2 <- c(0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0)
> cvec3 <- c(0, 0, 0, 0, 0, 1, 0, 0, 2, 0, 0, 0, 0)
> cvec4 <- c(0, 0, 0, 0, 0, 1, 0, 0, 3, 0, 0, 0, 0)
> cvec5 <- c(0, 0, 0, 0, 0, 1, 0, 0, 4, 0, 0, 0, 0)
> cvec6 <- c(0, 0, 0, 0, 0, 1, 0, 0, 5, 0, 0, 0, 0)
> cvec7 <- c(0, 0, 0, 0, 0, 1, 0, 0, 6, 0, 0, 0, 0)
> cvec8 <- c(0, 0, 0, 0, 0, 1, 0, 0, 6, 0, 0, 1, 0)
> cvec9 <- c(0, 0, 0, 0, 0, 1, 0, 0, 6, 0, 0, 2, 0)
> cvec10 <- c(0, 0, 0, 0, 0, 1, 0, 0, 6, 0, 0, 3, 0)
> cmat <- cbind(cvec1, cvec2, cvec3, cvec4, cvec5, cvec6, cvec7, cvec8, cvec9, cvec10)
> est <- t(cmat) %*% psi
> se <- sqrt(diag(t(cmat) %*% vpsi %*% cmat))
> zval <- est/se
> pval <- round(pnorm(-abs(zval)) * 2, 6)
> cbind(est, se, zval, pval)
                        se
cvec1    0.05990577 0.2208941   0.2711968 0.78624
cvec2   -1.80410099 0.1899213  -9.4992019 0.00000
cvec3   -3.66810775 0.1923457 -19.0703922 0.00000
cvec4   -5.53211451 0.2271001 -24.3598044 0.00000
cvec5   -7.39612127 0.2824941 -26.1815083 0.00000
cvec6   -9.26012803 0.3488298 -26.5462676 0.00000
cvec7  -11.12413479 0.4209660 -26.4252602 0.00000
cvec8   -9.93550928 0.3658028 -27.1608344 0.00000
cvec9   -8.74688377 0.4261681 -20.5244901 0.00000
cvec10  -7.55825826 0.5662465 -13.3479996 0.00000
```

ในวันที่ 6-14 ซึ่งอยู่ในช่วงบำบัด
หรือช่วงติดตาม พบว่ากลุ่มบำบัด
มีความเครียดน้อยกว่ากลุ่มควบคุม

ความแปรปรวนค่าคงเหลือไม่เท่ากัน

- โมเดลเส้นโค้งพัฒนาการจะมีสมมติฐานทางสถิติ (Assumption) ว่าความแปรปรวนของค่าคงเหลือเท่ากัน ในทุกๆ ระดับของเวลา (Homoscedasticity) กล่าวคือ e_{ij} มีความแปรปรวนเท่ากันในทุกค่าของ *i* และ *j* คือ σ^2
- แต่ในบางครั้ง σ^2 อาจเปลี่ยนแปลงไปตามเวลา เช่น คำศัพท์ที่ทารกพูดได้ ตอนแรกเกิดอาจมีความแตกต่างน้อย เช่น 0-3 คำ แต่พอเข้าวัยเตาะแตะ บางคนก็ยังพูดไม่ได้ บางคนกลับพูดได้เยอะมาก เช่น ช่วงอาจถึง 0-40 คำ ซึ่งแสดงให้เห็นว่าความแปรปรวนของคำศัพท์พูด เพิ่มขึ้นตามวัยด้วย

ความแปรปรวนค่าคงเหลือไม่เท่ากัน

- ในการวิเคราะห์พหุระดับ (หรือการวิเคราะห์ถดถอย) สามารถสร้างโมเดลที่ความแปรปรวนค่าคงเหลือขึ้นอยู่กับค่าของตัวแปรอิสระได้ โดยสามารถแบ่งออกได้หลายรูปแบบ
- ให้ X เป็นตัวแปรอิสระระดับใดก็ได้ ซึ่งในบทนี้จะหมายถึงตัวแปรเวลา เราสามารถสร้างโมเดลที่ความแปรปรวนค่าคงเหลือระดับที่ 1 ให้อยู่ในรูปแบบใดรูปแบบหนึ่ง ดังต่อไปนี้
  - ค่าความแปรปรวนค่าคงเหลือคงที่ (Homoscedasticity): $\mathrm{Var}(e_{ij}) = \sigma^2$
  - ค่าความแปรปรวนค่าคงเหลือไม่เท่ากัน (Heteroscedasticity): $\mathrm{Var}(e_{ij}) = f(X)$

ความแปรปรวนค่าคงเหลือไม่เท่ากัน

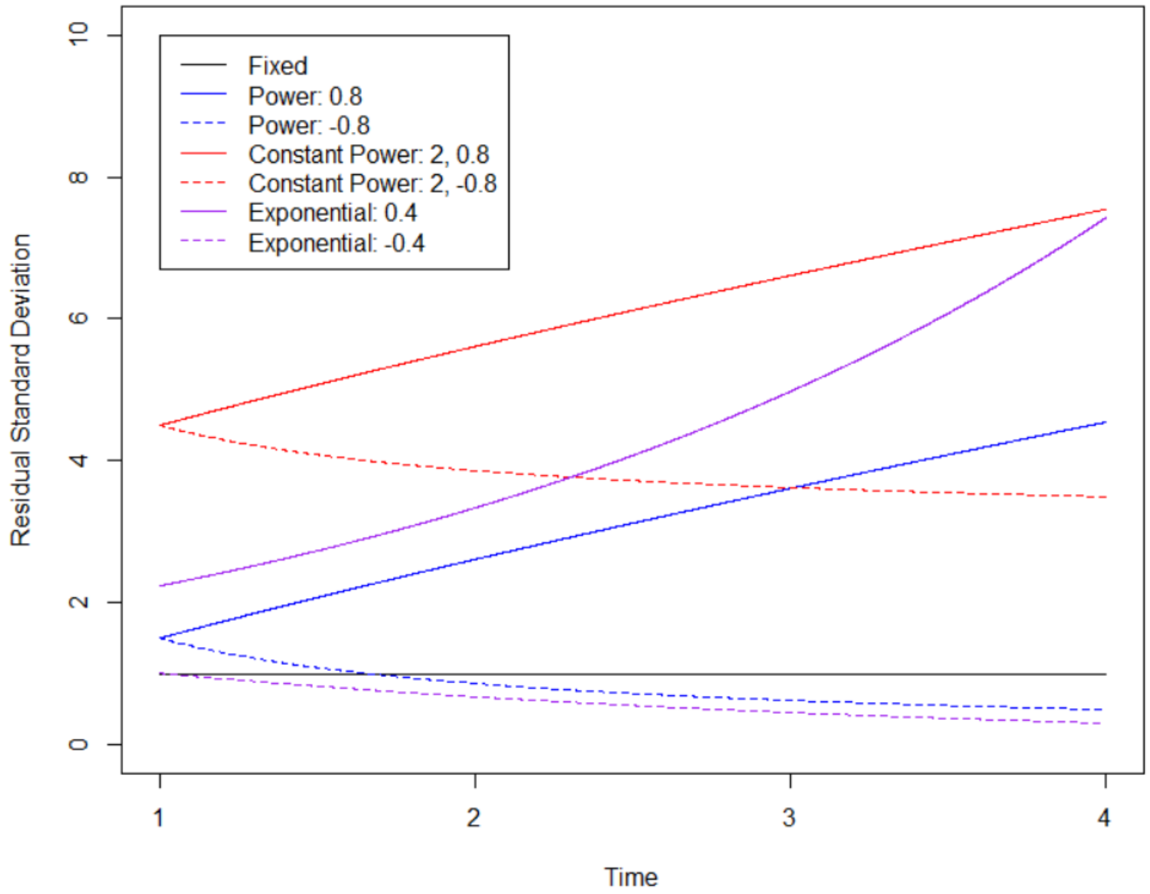
- รูปแบบของฟังก์ชั่นที่มีในโปรแกรม
  - ความแปรปรวนขึ้นอยู่กับค่าของ *X* กล่าวคือ $\mathrm{Var}(e_{ij}) = \sigma_x^2$ เช่น ความแปรปรวนของค่าคงเหลือที่วัดในปีที่ 1, 2, 3, และ 4 แตกต่างกัน
    - สำหรับวิธีการนี้ หากใช้กับตัวแปรเวลาจะเหมาะสมกับเวลาที่มีค่าจำกัด ถ้าค่าของเวลาของแต่ละคนแตกต่างกัน (Individually-varying time) จะไม่สามารถใช้ได้ เช่น คนแรกวัดวันที่ 0, 25, และ 55, คนที่สองวัดวันที่ 0, 17, และ 38, คนที่สามวัดวันที่ 0, 34, และ 42
  - ความแปรปรวนเปลี่ยนแปลงตามฟังก์ชั่นยกกำลัง (Power Function): $\mathrm{Var}(e_{ij}) = \sigma^2 \cdot |x|^{2k}$ โดย
    - $k < 0$ ยิ่ง *X* เพิ่ม ความแปรปรวนยิ่งลดลง
    - $k > 0$ ยิ่ง *X* เพิ่ม ความแปรปรวนยิ่งมากขึ้น

ความแปรปรวนค่าคงเหลือไม่เท่ากัน

- รูปแบบของฟังก์ชั่นที่มีในโปรแกรม
  - ความแปรปรวนเปลี่ยนแปลงตามฟังก์ชั่นยกกำลังบวกค่าคงที่ (Constant Power Function): $\mathrm{Var}\left(e_{ij}\right) = \sigma^2 \cdot \left(c + |x|^k\right)^2$
    - เหมือนฟังก์ชั่นยกกำลัง ค่า *c* เป็นเพียงตัวเลื่อนกราฟขึ้นลงตามแนวดิ่ง
  - ความแปรปรวนเปลี่ยนแปลงตามฟังก์ชั่น Exponential: $\mathrm{Var}\left(e_{ij}\right) = \sigma^2 \cdot e^{2kx}$ โดย
    - การเปลี่ยนแปลงเป็นรูปแบบ Exponential Function: $Y = \exp(kX)$
    - $k < 0$ ยิ่ง X เพิ่ม ความแปรปรวนยิ่งลดลง
    - $k > 0$ ยิ่ง X เพิ่ม ความแปรปรวนยิ่งมากขึ้น

ความแปรปรวนค่าคงเหลือไม่เท่ากัน

- ตรวจสอบการเปลี่ยนแปลง SD ของค่าคงเหลือ ให้ X = 1, 2, 3, 4
- Fixed: $SD_e = 1.5$
- Power: $SD_e = 1.5 \cdot |X|^k$
- Constant Power:
  $SD_e = 1.5 \cdot \left(c + |X|^k\right)$
- Exponential: $SD_e = e^{kX}$

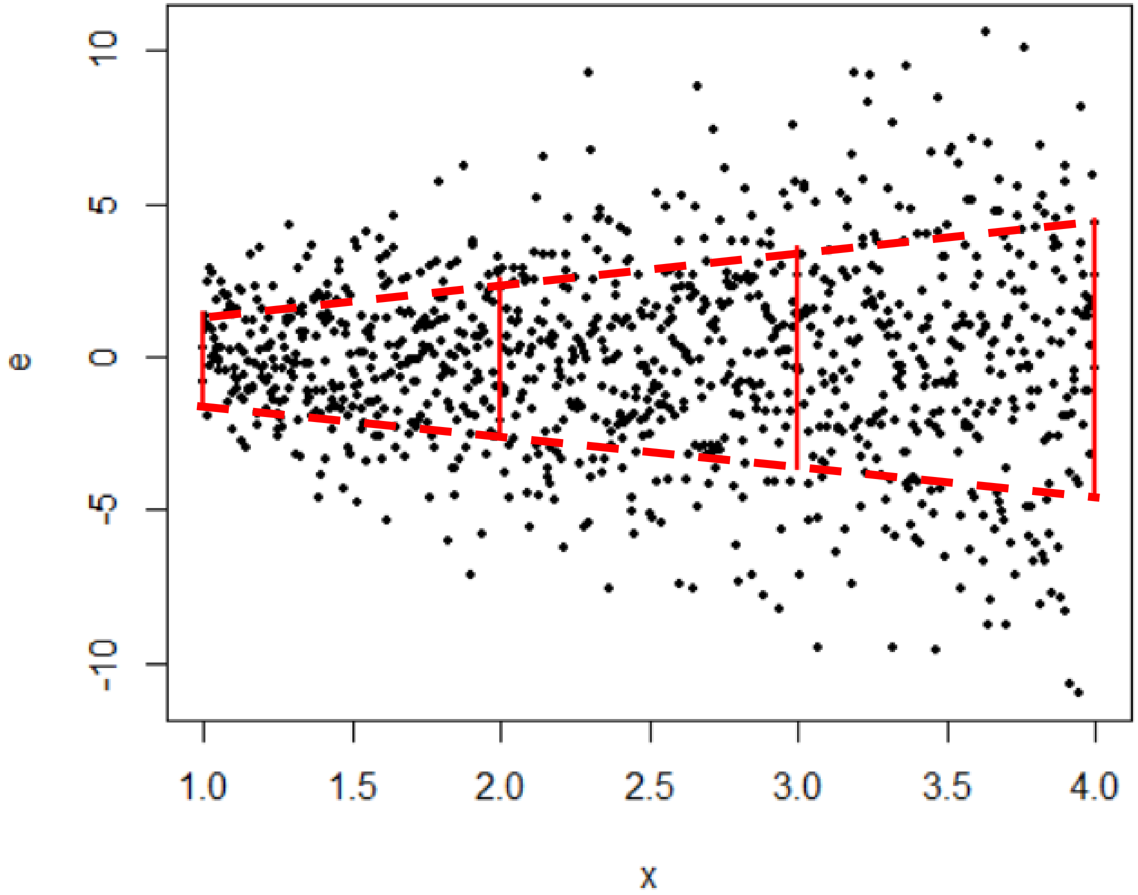


ความแปรปรวนค่าคงเหลือไม่เท่ากัน

- ตัวอย่างข้อมูลฟังก์ชั่น Power ที่ $SD_e = 1.5 \cdot |X|^{0.8}$

เส้นสีแดงมีความกว้าง

± 1 *SD* ในแต่ละช่วงเวลา



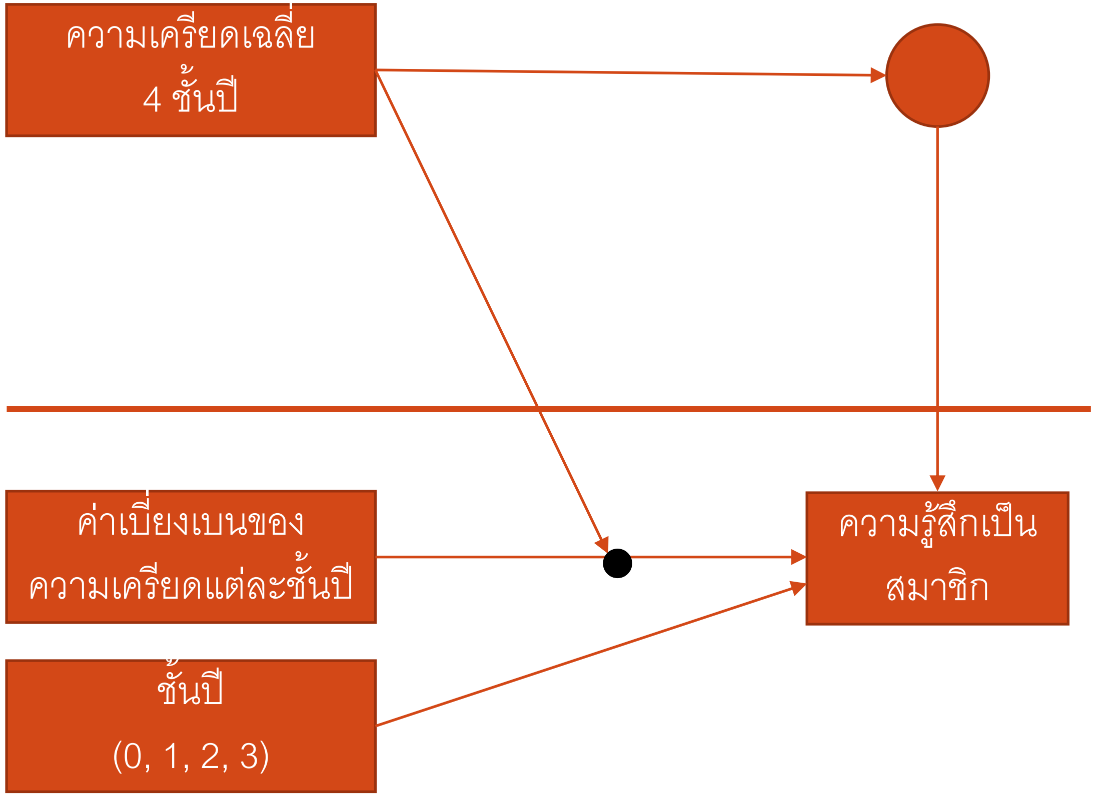
ความแปรปรวนค่าคงเหลือไม่เท่ากัน

- ในกรณีที่ X = 1, 2, 3, 4 หากกำหนดให้ค่าความแปรปรวนขึ้นอยู่กับค่าของ X จะหมายความว่า $\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2 \neq \sigma_4^2$ จากเดิมประมาณค่าพารามิเตอร์ตัวเดียว ตอนนี้ประมาณค่าพารามิเตอร์ 4 ค่า *df* จะลดลง 3
- ส่วนค่า *df* ที่ลดลงค่าฟังก์ชั่นความแปรปรวนโมเดลอื่น ขึ้นอยู่กับค่าพารามิเตอร์ที่ทำนายเพิ่ม เช่น ฟังก์ชั่นกำลัง จะทำให้ *df* ลดลง 1
- สามารถเปรียบเทียบโมเดลความแปรปรวนเท่าและแตกต่างกันได้ ด้วย Likelihood Ratio Test
- รายละเอียดเพิ่มเติมที่ https://data.library.virginia.edu/modeling-non-constant-variance/

ความแปรปรวนค่าคงเหลือไม่เท่ากัน

ทำนายความรู้สึกเป็นสมาชิก
ของนิสิตจากคณะแห่งหนึ่ง

เก็บข้อมูลจากนิสิต 850 คน เป็นนิสิตรุ่น 2550 ทั้งหมด 426 คน และนิสิตรุ่น 2551 จำนวน 424 คน เก็บข้อมูลระยะยาวตลอดเวลา 4 ปี



```
> dat5 <-read.table("lecture8ex2.csv", sep=",", header=TRUE)
> dat5$timec <- dat5$time - 1
> dat5$avestress <- ave(dat5$stress, dat5$pid)
> dat5$diffstress <- dat5$stress - dat5$avestress
> library(lme4)
> out5 <- lmer(mem ~ 1 + timec + diffstress + avestress + (1 + timec|pid), data=dat5, REML=FALSE)
> summary(out5)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula:
mem ~ 1 + timec + diffstress + avestress + (1 + timec | pid)
   Data: dat5

     AIC       BIC    logLik deviance df.resid
 25552.9   25602.0  -12768.5  25536.9     3392

Scaled residuals:
     Min       1Q   Median       3Q      Max
-3.10411 -0.48077  0.00681  0.49606  2.61723

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pid      (Intercept) 234.68   15.319
          timec        35.29    5.941   -0.15
 Residual              25.92    5.091
Number of obs: 3400, groups:  pid, 850

Fixed effects:
            Estimate Std. Error t value
(Intercept) 56.78964    2.57724  22.035
timec       -1.88457    0.21823  -8.636
diffstress   0.40878    0.01212  33.730
avestress    0.10050    0.05011   2.006
```

โมเดลเดิมที่กล่าวถึงในบทที่แล้ว ตอนนี้ ค่าความแปรปรวนคงเหลือระดับที่ 1 เท่ากันทุกช่วงเวลา เท่ากับ $\sigma^2 = 25.92$ โดยส่วนเบี่ยงเบนมาตรฐานคือ $\sigma = 5.09$

ความแปรปรวนค่าคงเหลือไม่เท่ากัน

- ในการวิเคราะห์โมเดลความแปรปรวนค่าคงเหลือไม่เท่ากัน (Heteroscedastic Model) เราจะใช้ `nlme` package ซึ่ง Doug Bates ผู้พัฒนา `lme4` เคยพัฒนามาก่อนที่ย้ายมาพัฒนา `lme4`
- `lme4` จะมีการวิเคราะห์ที่รวดเร็วกว่า เพราะโครงสร้างโปรแกรมการคำนวณออกแบบดีกว่าเดิม แต่ไม่ได้พัฒนาให้ครอบคลุมทุกตัวเลือกที่ `nlme` ทำได้

ความแปรปรวนค่าคงเหลือไม่เท่ากัน

- วิเคราะห์โมเดลเดิมใน `nlme`

เรียก package `nlme`

คำสั่ง Fixed Effect

```
> library(nlme)
> out5a <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit)
```

แยกเขียน Random Effect

เลือกได้ระหว่าง ML และ REML โดยค่าเริ่มต้นคือ REML แต่ผมเลือกใช้ ML

ค่าเริ่มต้น คือ `na.fail` ที่เมื่อเจอค่าสูญหายจะยกเลิกการวิเคราะห์ทันที ในที่นี้เลือก `na.omit` เพื่อลบแถวที่มีค่าสูญหายออก (listwise deletion)

```
> library(nlme)
> out5a <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit)
> summary(out5a)
Linear mixed-effects model fit by maximum likelihood
  Data: dat5
       AIC    BIC    logLik
  25552.95 25602 -12768.47

Random effects:
 Formula: ~1 + timec | pid
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 15.319143 (Intr)
timec        5.940565 -0.152
Residual     5.090837

Fixed effects:  mem ~ 1 + timec + diffstress + avestress
               Value Std.Error   DF  t-value p-value
(Intercept) 56.78964 2.5787499 2548 22.02216  0.0000
timec       -1.88457 0.2183537 2548 -8.63082  0.0000
diffstress   0.40878 0.0121265 2548 33.70962  0.0000
avestress    0.10050 0.0501375  848  2.00458  0.0453
 Correlation:
           (Intr) timec  dffstr
timec      -0.045
diffstress -0.004  0.011
avestress  -0.977  0.000  0.003

Standardized Within-Group Residuals:
         Min           Q1          Med          Q3          Max
-3.104097747 -0.480769826  0.006814801  0.496061463  2.617219474

Number of Observations: 3400
Number of Groups: 850
```

ผลการวิเคราะห์จะเหมือนกับ lme4

ค่าความเหมาะสมของโมเดล

ค่าส่วนเบี่ยงเบนมาตรฐาน และสหสัมพันธ์ของอิทธิพลสุ่ม

ค่าสถิติของอิทธิพลคงที่ ซึ่งในที่นี้ให้ *p*-value มาด้วย ต่างกับ lme4 ที่ไม่ให้ *p*-value มา

ค่าความสัมพันธ์ระหว่างอิทธิพลคงที่ แทบไม่ได้ใช้แปลความหมาย

ค่าสถิติของ $e_{ij}/\sqrt{\sigma^2}$

```
> out5b <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit,
+              weights=varIdent(form = ~1|timec))
> summary(out5b)
Linear mixed-effects model fit by maximum likelihood
  Data: dat5
      AIC      BIC    logLik
  25510.7 25578.15 -12744.35

Random effects:
 Formula: ~1 + timec | pid
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 15.332929 (Intr)
timec        5.888771 -0.154
Residual     6.527770

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | timec
 Parameter estimates:
        0         1         2         3
1.0000000 0.6696639 0.9035288 0.3313774
Fixed effects:  mem ~ 1 + timec + diffstress + avestre
               Value Std.Error   DF  t-value p-value
(Intercept) 56.91609 2.5920724 2548 21.95776  0.0000
timec       -1.78621 0.2110655 2548 -8.46280  0.0000
diffstress   0.40682 0.0119806 2548 33.95679  0.0000
avestress    0.09579 0.0503951  848  1.90079  0.0577
```

เพิ่มบรรทัดนี้มา เพื่อทำ Heteroscedastic Model ที่ค่าความแปรปรวนของค่าคงเหลือในกลุ่มของ timec เดียวกัน จะมีค่าเท่ากัน

σ มีค่าเท่ากับ 6.528

ค่าคงเหลือแต่ละชั้นปีมี SD ดังนี้

- $\sigma_1 = 6.528 \times 1 = 6.528$
- $\sigma_2 = 6.528 \times 0.670 = 4.371$
- $\sigma_3 = 6.528 \times 0.904 = 5.898$
- $\sigma_4 = 6.528 \times 0.331 = 2.163$

```
> anova(out5a, out5b)
      Model df      AIC      BIC    logLik   Test L.Ratio p-value
out5a     1  8 25552.95 25602.00 -12768.47
out5b     2 11 25510.70 25578.15 -12744.35 1 vs 2 48.2462  <.0001
```

ค่าแปรปรวนคงเหลือของทั้ง 4 เวลาแตกต่างกันอย่างมีนัยสำคัญ จึงใช้โมเดลค่าคงเหลือมีความแปรปรวนแตกต่างกัน (Heteroscedastic Model) ทั้ง 4 เวลา

```
> out5bsigma <- as.numeric(VarCorr(out5b)[3, 2])
> out5bsigmatime <- out5bsigma * coef(out5b$modelStruct$varStruct, uncons = FALSE)
> c(out5bsigma, out5bsigmatime)
                1        2        3
6.527770 4.371412 5.898028 2.163155
```

คำนวณค่าส่วนเบี่ยงเบนมาตรฐานของแต่ละช่วงเวลา

```
> out5c <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit,
+              weights=varExp(form = ~time))
> summary(out5c)
Linear mixed-effects model fit by maximum likelihood
 Data: dat5
       AIC      BIC    logLik
  25552.81 25607.99 -12767.4

Random effects:
 Formula: ~1 + timec | pid
 Structure: General positive-definite, Log-Cholesky param
            StdDev    Corr
(Intercept) 15.271202 (Intr)
timec        5.899540 -0.144
Residual     5.560076

Variance function:
 Structure: Exponential of variance covariate
 Formula: ~time
 Parameter estimates:
      expon
-0.03465453
```

เพิ่มบรรทัดนี้มา เพื่อทำ Heteroscedastic Model ที่ค่าความแปรปรวนของค่าคงเหลือเป็นฟังก์ชั่น Exponential ของ `time`

$$\sigma_t = \sigma \cdot e^{kt} = 5.37 \cdot e^{-0.035t}$$

```
> out5csigma <- as.numeric(VarCorr(out5c)[3, 2])
> expvalue <- coef(out5c$modelStruct$varStruct, uncons = FALSE)
> out5csigma * exp(expvalue * (1:4))
[1] 5.370695 5.187764 5.011064 4.840382
```

ความแปรปรวนแต่ละช่วงเวลา

```
> anova(out5a, out5c)
      Model df      AIC      BIC    logLik   Test L.Ratio p-value
out5a     1  8 25552.95 25602.00 -12768.47
out5c     2  9 25552.81 25607.99 -12767.40 1 vs 2 2.14082  0.1434
```

ไม่แตกต่างจาก homoscedastic model อย่างมีนัยสำคัญ

```
> out5d <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit,
+              weights=varPower(form = ~time))
> summary(out5d)
Linear mixed-effects model fit by maximum likelihood
 Data: dat5
       AIC      BIC    logLik
  25551.99 25607.17 -12766.99

Random effects:
 Formula: ~1 + timec | pid
 Structure: General positive-definite, Log-Cholesky parame
             StdDev    Corr
(Intercept) 15.252170 (Intr)
timec        5.883418 -0.141
Residual     5.495261

Variance function:
 Structure: Power of variance covariate
 Formula: ~time
 Parameter estimates:
      power
-0.08991353
```

เพิ่มบรรทัดนี้มา เพื่อทำ Heteroscedastic Model ที่ค่าความแปรปรวนของค่าคงเหลือเป็นฟังก์ชั่น Power ของ `time`

$$\sigma_t = \sigma \cdot |t|^k = 5.495 \cdot |t|^{-0.09}$$

```
> out5dsigma <- as.numeric(VarCorr(out5d)[3, 2])
> powervalue <- coef(out5d$modelStruct$varStruct, uncons = FALSE)
> out5dsigma * (1:4)^powervalue
[1] 5.495261 5.163232 4.978387 4.851265
```

ความแปรปรวนแต่ละช่วงเวลา

```
> anova(out5a, out5d)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
out5a     1  8 25552.95 25602.00 -12768.47
out5d     2  9 25551.99 25607.17 -12767.00 1 vs 2 2.958496  0.0854
```

ไม่แตกต่างจาก homoscedastic model อย่างมีนัยสำคัญ

```
> out5e <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit,
+              weights=varConstPower(form = ~time),
+              control = lmeControl(maxIter=200, msMaxIter = 200))
> summary(out5e)
Linear mixed-effects model fit by maximum likelihood
 Data: dat5
       AIC      BIC   logLik
  25547.19 25608.51 -12763.6

Random effects:
 Formula: ~1 + timec | pid
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 15.192037 (Intr)
timec        5.813103 -0.126
Residual     1.157416

Variance function:
 Structure: Constant plus power of variance covariate
 Formula: ~time
 Parameter estimates:
     const      power
  4.246259 -22.457879
```

เพิ่มบรรทัดนี้มา เพื่อทำ Heteroscedastic Model ที่ค่าความแปรปรวนของค่าคงเหลือเป็นฟังก์ชั่น Constant Power ของ time

เพิ่มจำนวนรอบในการหาค่าพารามิเตอร์ของ ML เพื่อให้เจอ Convergent Solution

$$\sigma_t = \sigma \cdot \left(c + |t|^k\right) = 1.157 \cdot \left(4.246 + t^{-22.458}\right)$$

```
> out5esigma <- as.numeric(VarCorr(out5e)[3, 2])
> paramvar <- coef(out5e$modelStruct$varStruct, uncons = FALSE)
> out5esigma * (paramvar[1] + (1:4)^paramvar[2])
[1] 6.072104 4.914688 4.914688 4.914688
```

ความแปรปรวนแต่ละช่วงเวลา

```
> anova(out5a, out5e)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
out5a     1  8 25552.95 25602.00 -12768.47
out5e     2 10 25547.19 25608.51 -12763.60 1 vs 2 9.753751  0.0076
```

แตกต่างจาก homoscedastic model อย่างมีนัยสำคัญ

โมเดล	ปีที่ 1	ปีที่ 2	ปีที่ 3	ปีที่ 4
a) คงที่	5.091	5.091	5.091	5.091
b) ต่างกันแต่ละช่วงเวลา	6.528	4.371	5.898	2.163
c) Exponential	5.371	5.188	5.011	4.840
d) Power	5.495	5.163	4.978	4.851
e) Constant Power	6.072	4.915	4.915	4.915

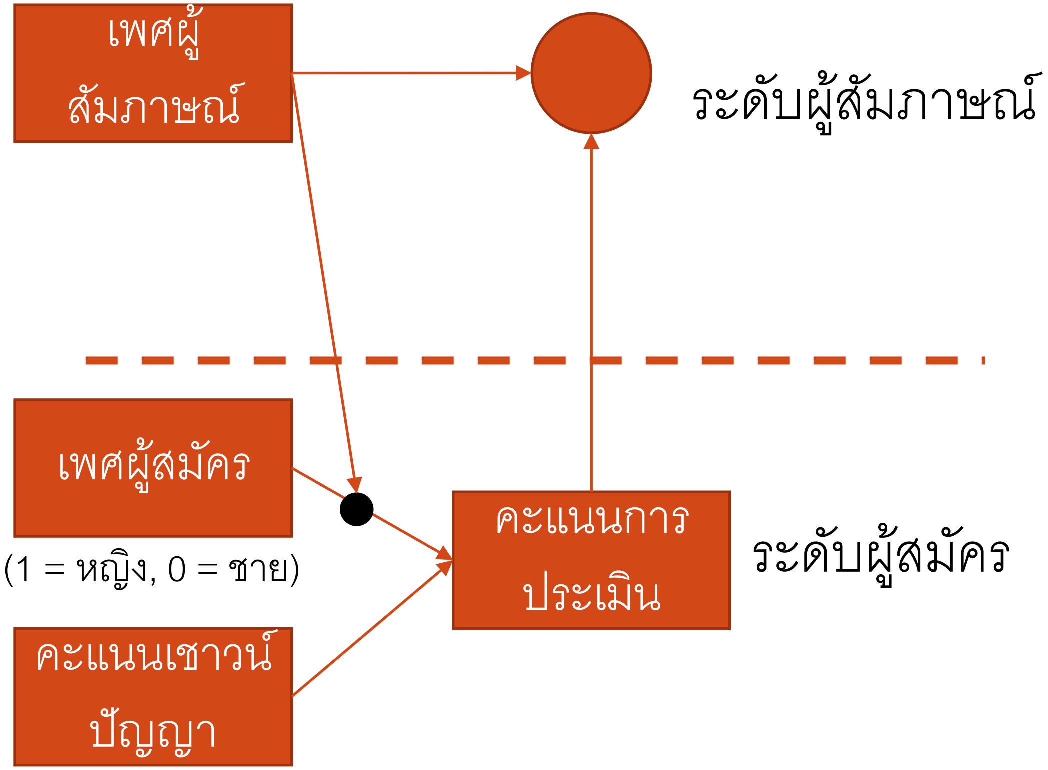
โมเดล b และ e ดีกว่าโมเดล a ทั้งคู่ เปรียบเทียบโมเดล b และ e ด้วย Likelihood Ratio Test เนื่องจากโมเดลทั้งสองซ้อนกัน โดยความแปรปรวนของโมเดล b ทำนายด้วย 4 พารามิเตอร์ แต่โมเดล c ทำนายด้วย 3 พารามิเตอร์

```
> anova(out5e, out5b)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
out5e     1 10 25547.19 25608.51 -12763.60
out5b     2 11 25510.70 25578.15 -12744.35 1 vs 2 38.49245  <.0001
```

โมเดล b เป็นโมเดลที่อธิบายความแปรปรวนได้ดีที่สุด

ความแปรปรวนค่าคงเหลือไม่เท่ากัน

ทดสอบโมเดลที่ความแปรปรวนของค่าคงเหลือ
แตกต่างกันระหว่างเพศผู้สัมภาษณ์และ
เพศของผู้สมัคร



Homoscedastic Model

```
> dat6 <-read.table("lecture4ex2.csv", sep=",", header=TRUE)
> out6a <- lme(score ~ 1 + ersex + eesex + eesex:ersex + I(iq - 100),
+              random = ~1 + eesex|erid,
+              data=dat6, method="ML", na.action=na.omit)
> summary(out6a)
Linear mixed-effects model fit by maximum likelihood
  Data: dat6
       AIC      BIC   logLik
  61640.99 61705.88 -30811.5

Random effects:
 Formula: ~1 + eesex | erid
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 4.4849853 (Intr)
eesex       0.8932369 0.125
Residual    4.6652498

Fixed effects:  score ~ 1 + ersex + eesex + eesex:ersex + I(iq - 100)
               Value  Std.Error   DF  t-value p-value
(Intercept) 75.53870 0.22127123 8997 341.3851  0.0000
ersex        0.00989 0.31292407  998   0.0316  0.9748
eesex       -0.41323 0.13790183 8997  -2.9966  0.0027
I(iq - 100)  0.02942 0.00329961 8997   8.9155  0.0000
ersex:eesex  1.07365 0.19502919 8997   5.5051  0.0000
```

```
> out6b <- lme(score ~ 1 + ersex + eesex + eesex:ersex + I(iq - 100),
+              random = ~1 + eesex|erid,
+              data=dat6, method="ML", na.action=na.omit,
+              weights=varIdent(form = ~1|eesex))
> summary(out6b)
Linear mixed-effects model fit by maximum likelihood
  Data: dat6
       AIC      BIC    logLik
  61642.39 61714.49 -30811.19

Random effects:
 Formula: ~1 + eesex | erid
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 4.4790343 (Intr)
eesex       0.8932343 0.138
Residual    4.6937412

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | eesex
 Parameter estimates:
        0         1
1.0000000 0.9878226

 Fixed effects:  score ~ 1 + ersex + eesex + eesex:ersex + I(iq - 100)
                Value  Std.Error   DF  t-value p-value
(Intercept) 75.53870 0.22127116 8997 341.3852  0.0000
ersex        0.00989 0.31292398  998   0.0316  0.9748
eesex       -0.41323 0.13790180 8997  -2.9966  0.0027
I(iq - 100)  0.02941 0.00329918 8997   8.9143  0.0000
ersex:eesex  1.07365 0.19502915 8997   5.5051  0.0000
```

ความแปรปรวนแตกต่างกันระหว่างเพศผู้สมัคร

ผู้สมัคร

เพศชาย	เพศหญิง
4.694	4.694 x 0.988 = 4.638

```
> anova(out6a, out6b)
      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
out6a     1  9 61640.99 61705.88 -30811.49
out6b     2 10 61642.39 61714.49 -30811.19 1 vs 2 0.6004415  0.4384
```

ความแปรปรวนระหว่างเพศผู้สมัครไม่ต่างกันอย่างมีนัยสำคัญ

```
> out6c <- lme(score ~ 1 + ersex + eesex + eesex:ersex + I(iq - 100),
+              random = ~1 + eesex|erid,
+              data=dat6, method="ML", na.action=na.omit,
+              weights=varIdent(form = ~1|ersex))
> summary(out6c)
Linear mixed-effects model fit by maximum likelihood
  Data: dat6
       AIC      BIC    logLik
  61642.85 61714.95 -30811.43

Random effects:
 Formula: ~1 + eesex | erid
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 4.4850480 (Intr)
eesex       0.8929552 0.125
Residual    4.6783279

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | ersex
 Parameter estimates:
        0         1
1.0000000 0.9944001

Fixed effects:  score ~ 1 + ersex + eesex + eesex:ersex + I(iq - 100)
               Value  Std.Error   DF  t-value p-value
(Intercept) 75.53871 0.22138425 8997 341.2108  0.0000
ersex        0.00988 0.31292764  998   0.0316  0.9748
eesex       -0.41324 0.13825236 8997  -2.9890  0.0028
I(iq - 100)  0.02939 0.00329958 8997   8.9061  0.0000
ersex:eesex  1.07366 0.19502394 8997   5.5053  0.0000

> anova(out6a, out6c)
      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
out6a     1  9 61640.99 61705.88 -30811.49
out6c     2 10 61642.85 61714.95 -30811.42 1 vs 2 0.1395039  0.7088
```

ความแปรปรวนแตกต่างกันระหว่างเพศผู้สัมภาษณ์

ผู้สัมภาษณ์

เพศชาย	เพศหญิง
4.678	4.678 x 0.994 = 4.650

ความแปรปรวนระหว่างเพศผู้สัมภาษณ์ไม่ต่างกันอย่างมีนัยสำคัญ

```
> out6d <- lme(score ~ 1 + ersex + eesex + eesex:ersex + I(iq - 100),
+              random = ~1 + eesex|erid,
+              data=dat6, method="ML", na.action=na.omit,
+              weights=varIdent(form = ~1|eesex * ersex))
> summary(out6d)
Linear mixed-effects model fit by maximum likelihood
  Data: dat6
       AIC      BIC    logLik
  61646.21 61732.73 -30811.11

Random effects:
 Formula: ~1 + eesex | erid
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 4.4791338 (Intr)
eesex       0.8928087 0.139
Residual    4.7142126

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | eesex * ersex
 Parameter estimates:
      0*0       1*0       0*1       1*1
1.0000000 0.9847240 0.9913071 0.9823362

Fixed effects:  score ~ 1 + ersex + eesex + eesex:ersex + I(iq - 100)
               Value  Std.Error   DF   t-value p-value
(Intercept) 75.53871 0.22144929 8997 341.1106  0.0000
ersex        0.00988 0.31292999  998   0.0316  0.9748
eesex       -0.41324 0.13825089 8997  -2.9890  0.0028
I(iq - 100)  0.02938 0.00329901 8997   8.9058  0.0000
ersex:eesex  1.07367 0.19502165 8997   5.5054  0.0000

> anova(out6a, out6d)
      Model df      AIC      BIC    logLik   Test   L.Ratio p-value
out6a     1  9 61640.99 61705.88 -30811.49
out6d     2 12 61646.21 61732.73 -30811.10 1 vs 2 0.7796672  0.8543
```

ความแปรปรวนแตกต่างกันระหว่างทั้งเพศผู้สัมภาษณ์และเพศผู้สมัคร

ผู้สัมภาษณ์

ผู้สมัคร

	เพศชาย	เพศหญิง
เพศชาย	4.714	4.714 x 0.985 = 4.643
เพศหญิง	4.714 x 0.991 = 4.672	4.714 x 0.982 = 4.629

ความแปรปรวนทั้งสี่กลุ่ม
ไม่ต่างกันอย่างมีนัยสำคัญ

ความสัมพันธ์ในตัว (Autocorrelation)

- ให้ Y_{tj} เป็นค่าของตัวแปรตามเวลาที่ *t* ในคนที่ *j* สมมติว่าไม่มีการเปลี่ยนแปลงค่าเฉลี่ยระหว่างช่วงเวลา ทำให้ Y_{tj} มีความสัมพันธ์ตามโมเดลพหุระดับดังต่อไปนี้

$$Y_{tj} = \gamma_{00} + u_{0j} + e_{tj}$$ (ไม่มี T_{tj} เพราะไม่มีการเปลี่ยนแปลงค่าเฉลี่ย)

- หากโมเดลนี้เป็นจริง ความสัมพันธ์ระหว่างตัวแปรเป็นดังนี้

$$Cov\left(Y_{tj}, Y_{t'j}\right) = \begin{bmatrix} \tau_{00} + \sigma^2 & \tau_{00} & \cdots & \tau_{00} \\ \tau_{00} & \tau_{00} + \sigma^2 & \cdots & \tau_{00} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{00} & \tau_{00} & \cdots & \tau_{00} + \sigma^2 \end{bmatrix}$$

ความสัมพันธ์ในตัว (Autocorrelation)

- ดังนั้น ความแปรปรวนร่วมระหว่างคะแนนของคนใดคนหนึ่งใน 2 ช่วงเวลา เช่น เวลาที่ 1 และ 2 หรือเวลาที่ 1 และ 4 เท่ากับ τ_{00}
- ความแปรปรวนของ *Y* จะมีค่าเท่ากันในทุกคนและทุกช่วงเวลา คือ $\tau_{00} + \sigma^2$
- รูปแบบเมทริกซ์ที่สมาชิกแนวทแยงเท่ากัน และสมาชิกนอกแนวทแยงเท่ากัน จะเรียกว่าเมทริกซ์สมมาตรประกอบ (Compound Symmetry) กล่าวคือ มีค่าความแปรปรวนเท่ากันเพียงค่าเดียว และมีค่าความแปรปรวนร่วมเท่ากันเพียงค่าเดียว

ความสัมพันธ์ในตัว (Autocorrelation)

- จากสมการข้างต้น ความสัมพันธ์ระหว่างค่าคงเหลือระดับที่ 1 จะมีรูปแบบดังนี้

$$Cov\left(e_{tj}, e_{t'j}\right) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_{T \times T}$$

- หมายความว่า ความแปรปรวนร่วมระหว่างคะแนนในคนเดียวกัน จะเกิดจากการเป็นสมาชิกคนเดียวกันเท่านั้น ซึ่งเท่ากับ $\boldsymbol{\tau_{00}}$

ความสัมพันธ์ในตัว (Autocorrelation)

- อย่างไรก็ตาม ในสถานการณ์จริง เรามักจะพบว่าค่าคงเหลือที่มาจากคนเดียวกันในเวลาที่ใกล้กัน มักจะมีความคล้ายคลึงกัน เช่น ในกรณีที่ไม่มีตัวแปรอิสระ $Cov(Y_{1j}, Y_{2j})$ ควรจะมากกว่า $Cov(Y_{1j}, Y_{4j})$
  - เช่น วัดอารมณ์ทุกๆ ชั่วโมง เวลาในชั่วโมงที่ติดกัน ควรจะมีอารมณ์คล้ายคลึงกัน ตอน 10:00 ผู้ร่วมการทดลองคนหนึ่งอารมณ์เสีย เวลา 11:00 น่าจะยังมีอารมณ์เสียค้างอยู่ คะแนนเวลา 10 และ 11 โมงควรมีความสัมพันธ์กัน และความสัมพันธ์กันนี้ควรมากกว่าเวลาที่ห่างกัน
- ปรากฏการณ์ดังกล่าวเรียกว่า ความสัมพันธ์ในตัว (Autocorrelation) ซึ่งสามารถทำโมเดลพหุระดับเพื่อตรวจสอบปรากฏการณ์ดังกล่าวได้ดังนี้

ความสัมพันธ์ในตัว (Autocorrelation)

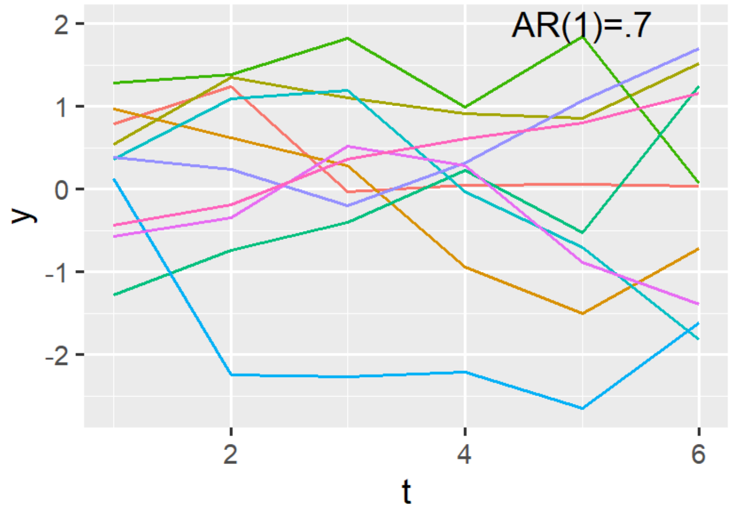
- โมเดลแบบง่ายที่สุดในการตรวจจับอิทธิพลระหว่างค่าคงเหลือ คือ โมเดลความสัมพันธ์ในตัวแบบง่าย (Simplex model) โดย

$$\left.\begin{aligned} Y_{tj} &= \gamma_{00} + u_{0j} + e_{tj} \\ e_{tj} &= \rho e_{t-1,j} + r_{tj} \end{aligned}\right\} \text{สมมติว่าไม่มีการเปลี่ยนแปลงค่าเฉลี่ย}$$

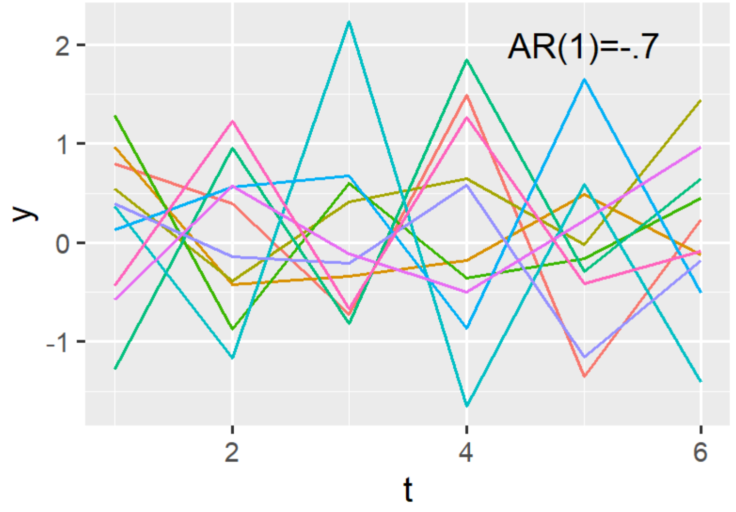
- e_t คือ ค่าคงเหลือในเวลาที่ t
- ρ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ในตัว (Autocorrelation; AR(1)) มีค่าตั้งแต่ -1 ถึง 1
- r_t คือ ค่าสุ่มที่เกิดขึ้นในเวลาที่ t มีความแปรปรวนเท่ากับ σ^2

ความสัมพันธ์ในตัว (Autocorrelation)

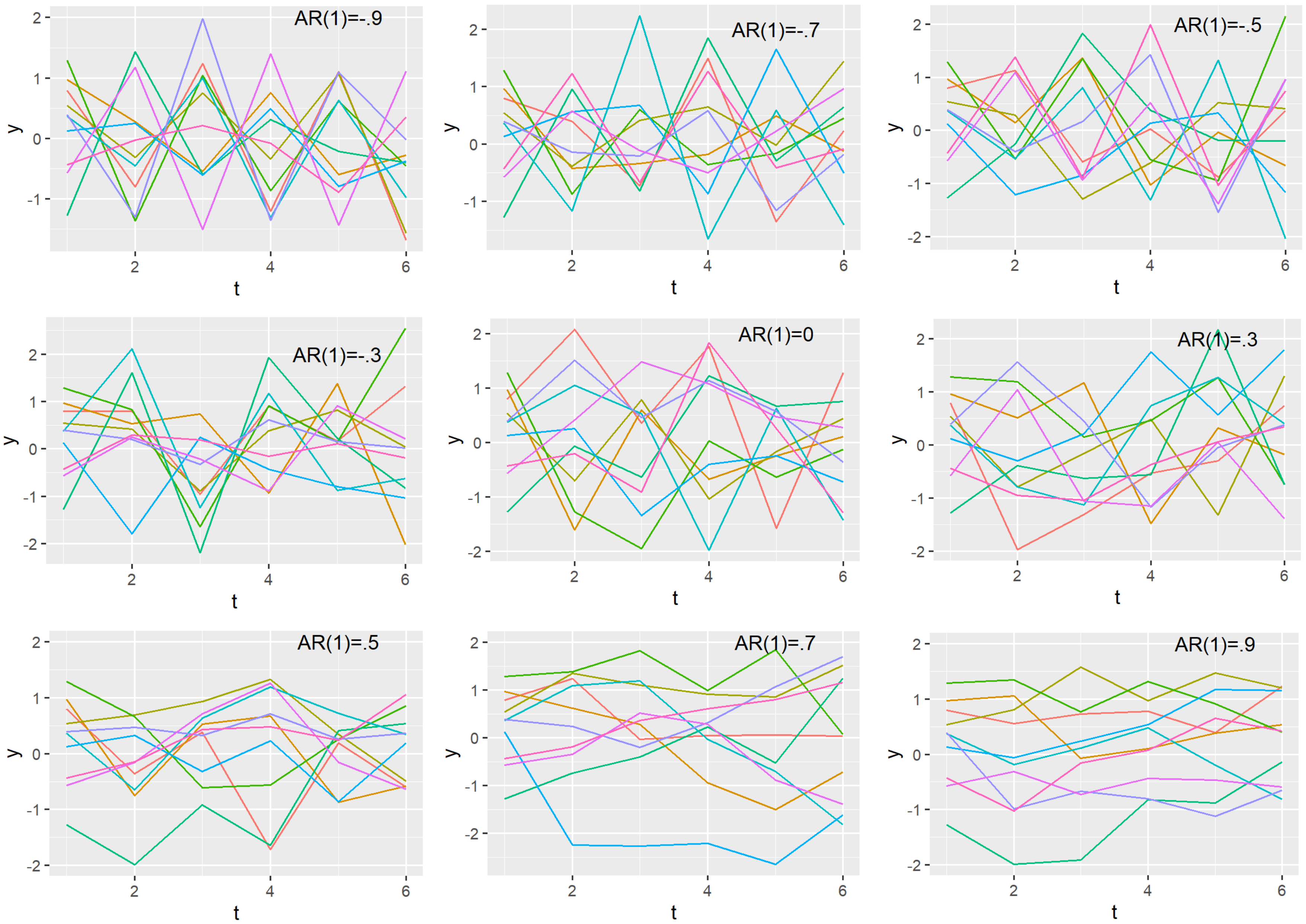
- หาก $\rho > 0$ ค่าที่อยู่ในเวลาใกล้กัน จะคล้ายคลึงกัน ยิ่งค่าสูง ยิ่งคล้ายกัน



- หาก $\rho < 0$ ค่าที่อยู่ในเวลาใกล้กัน จะมีค่าใกล้กัน แต่กระโดดไปฝั่งตรงข้ามของเส้น y = 0 ยิ่งขนาดสูง ยิ่งเห็นเป็นฟันปลา



สมการที่ใช้สร้างข้อมูล : $Y_{ij} = \gamma_{00} + e_{ij}$ $\quad e_{tj} = \rho e_{t-1,j} + r_{tj}$


AR(1)=-.9
AR(1)=-.7
AR(1)=-.5
AR(1)=-.3
AR(1)=0
AR(1)=.3
AR(1)=.5
AR(1)=.7
AR(1)=.9
y
t

จาก $e_{tj} = \rho e_{t-1,j} + r_{tj}$

คำนวณความแปรปรวนของค่าคงเหลือแต่ละช่วงเวลา

$$Var(e_{tj}) = Var(\rho e_{t-1,j}) + Var(r_{tj})$$

e และ *r* ไม่สัมพันธ์กัน

$$Var(e_{tj}) = \rho^2 Var(e_{t-1,j}) + \sigma^2$$

$$Var(e_{tj}) = \rho^2 Var(e_{tj}) + \sigma^2$$

ความแปรปรวนแต่ละช่วงเวลาเท่ากัน

$$(1-\rho^2)Var(e_{tj}) = \sigma^2$$

$$Var(e_{tj}) = \frac{\sigma^2}{1-\rho^2}$$

คำนวณความแปรปรวนร่วมของค่าคงเหลือในช่วงเวลาที่ติดกัน

$$Cov(e_{tj}, e_{t-1,j}) = Cov(\rho e_{t-1,j}, e_{t-1,j}) + Cov(r_{tj}, e_{t-1,j})$$

e และ *r* ไม่สัมพันธ์กัน
r ต่างช่วงเวลาไม่สัมพันธ์กัน

$$= \rho Var(e_{t-1,j})$$

$$= \frac{\rho\sigma^2}{1-\rho^2}$$

จาก $e_{tj} = \rho e_{t-1,j} + r_{tj} = \rho\left(\rho e_{t-2,j} + r_{t-1,j}\right) + r_{tj} = \rho^2 e_{t-2,j} + \rho r_{t-1,j} + r_{tj}$

คำนวณความแปรปรวนร่วมของค่าคงเหลือในช่วงเวลาที่ห่างกัน 2 ช่วง

$$Cov\left(e_{tj}, e_{t-2,j}\right) = Cov\left(\rho^2 e_{t-2,j}, e_{t-2,j}\right) + Cov\left(\rho r_{t-1,j}, e_{t-1,j}\right) + Cov\left(r_{tj}, e_{t-1,j}\right)$$

$$= \rho Var\left(e_{t-2,j}\right)$$

$$= \frac{\rho^2 \sigma^2}{1-\rho^2}$$

e และ *r* ไม่สัมพันธ์กัน

r ต่างช่วงเวลาไม่สัมพันธ์กัน

จากแนวทางการวิเคราะห์ด้านบน สามารถพิสูจน์ได้ว่า

$$Cov\left(e_{tj}, e_{t-k,j}\right) = \frac{\rho^k \sigma^2}{1-\rho^2}$$

ความสัมพันธ์ในตัว (Autocorrelation)

- ดังนั้น เมทริกซ์ความแปรปรวนร่วมเป็นดังนี้

$$Cov\left(e_{tj}, e_{t'j}\right) = \frac{\sigma^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}$$

$$Cov\left(Y_{tj}, Y_{t'j}\right) = Cov\left(e_{tj}, e_{t'j}\right) + \tau_{00}\mathbf{1}_{T\times T}$$

- จะเห็นว่าในโมเดลนี้ ยิ่งเวลาห่างความสัมพันธ์ยิ่งน้อยลง

ความสัมพันธ์ในตัว (Autocorrelation)

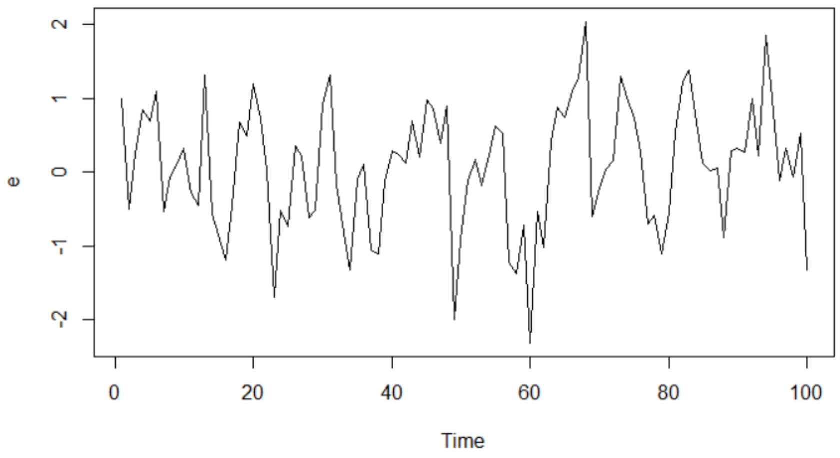
- นอกจากโมเดล AR(1) ยังมีโมเดล AR(2)

$$e_{tj} = \rho_1 e_{t-1,j} + \rho_2 e_{t-2,j} + r_{tj}$$

- โมเดลนี้ จะมีความเหมาะสมเมื่อรูปแบบตัวแปรจะออกมาเป็นแบบคลื่น เช่น การเปลี่ยนแปลงอารมณ์ทางบวก

$$e_t = .4e_{t-1} - .1e_{t-2} + r_t$$

$$r_t \sim N(0{,}0.71), e_1 = 1, e_2 = -0.5$$

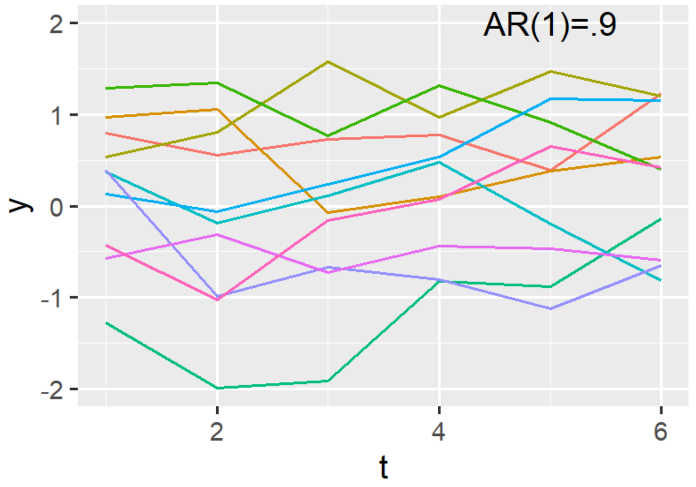


ความสัมพันธ์ในตัว (Autocorrelation)

- ข้อตกลงเบื้องต้นของโมเดลนี้ คือ โมเดลความสัมพันธ์ในตัว ยังเป็นแบบเดิมในช่วงเวลาทุกช่วงเวลา หรือที่เรียกว่า ความนิ่งของอนุกรมเวลา (Stationary of Time Series Model)

ความสัมพันธ์ในตัว (Autocorrelation)

- อย่างไรก็ตาม ข้อมูลที่เกิดจากโมเดลที่มีสหสัมพันธ์ในตัว อาจดูใกล้เคียงกับโมเดลอื่นๆ ในการวิเคราะห์พหุระดับ เช่น



โมเดลนี้ อาจมองว่าเป็นโมเดลสหสัมพันธ์ในตัวได้

หรืออาจมองเป็นโมเดลที่มีจุดตัดแกน *Y* แบบสุ่ม

(Random Intercept Model) ได้

กล่าวคือ ทั้งสองโมเดลอาจเหมาะสมกับข้อมูลทั้งคู่ และการเลือกโมเดลที่เหมาะสมควรเลือกจากทฤษฎี

ความสัมพันธ์ในตัว (Autocorrelation)

- ความสัมพันธ์ของตัวแปรตามในแต่ละช่วงเวลาในโมเดลที่มีความชันสุ่ม จะไม่ใช่เมทริกซ์สมมาตรประกอบ (Compound Symmetry) โดย

$$\mathrm{Var}\left(Y_{tj}\right) = \tau_{00} + \tau_{01}(T_t - T_0) + \tau_{11}(T_t - T_0) + \sigma^2$$

$$\mathrm{Cov}\left(Y_{tj}, Y_{t'j}\right) = \tau_{00} + \tau_{01}\left((T_t - T_0) + \left(T_{t'} - T_0\right)\right) + \tau_{11}(T_t - T_0)\left(T_{t'} - T_0\right)$$

- $T_t - T_0$ และ $T_{t'} - T_0$ คือ ระยะห่างของจุดเวลาที่สนใจกับจุดที่เวลาตั้งศูนย์กลางไว้

ความสัมพันธ์ในตัว (Autocorrelation)

- อย่างไรก็ตาม โมเดลที่มีความชันสุ่ม (Random Slope Model) ก็ทำให้ค่าของตัวแปรตาม มีความแปรปรวนเปลี่ยนแปลงไปตามเวลา และความสัมพันธ์ของตัวแปรตามระหว่างสองช่วงเวลา ก็ไม่ได้มีค่าคงที่ ขึ้นอยู่กับช่วงเวลาทั้งสองจุด
- ดังนั้น ข้อมูลที่เหมาะสมกับความสัมพันธ์ในตัว อาจจะเหมาะสมกับโมเดลที่มีความชันสุ่มด้วย กล่าวคือ โมเดลทั้งสองก็เป็นไปได้ทั้งคู่
- ในกรณีนี้ การเลือกโมเดลที่เหมาะสมควรเลือกจากทฤษฎี

ความสัมพันธ์ในตัว (Autocorrelation)

- Verbeke และ Lesaffre (1997) ได้ทดลองและพบว่าโมเดลพหุระดับที่ไม่ได้สนใจโครงสร้างของความสัมพันธ์ระหว่างตัวแปรตามในแต่ละช่วงเวลาอาจเป็นโมเดลที่ใช้ได้ การประมาณค่าพารามิเตอร์ในโมเดลถูกต้อง หากการละเมิดนั้น ไม่ใช่การละเมิดที่รุนแรง
- ส่วนตัวผม เชื่อว่าความสัมพันธ์ในตัว (Autocorrelation) เป็นปรากฏการณ์ที่มีจริง แต่มักไม่ใช่ค่าพารามิเตอร์ที่นักวิจัยสนใจ ในการวิเคราะห์พหุระดับ
- แนะนำให้วิเคราะห์อิทธิพลของตัวแปรเวลาและตัวแปรอิสระต่อตัวแปรตามก่อน หากได้โมเดลที่สมบูรณ์แล้ว ค่อยวิเคราะห์ความสัมพันธ์ในตัวเพิ่มเติม

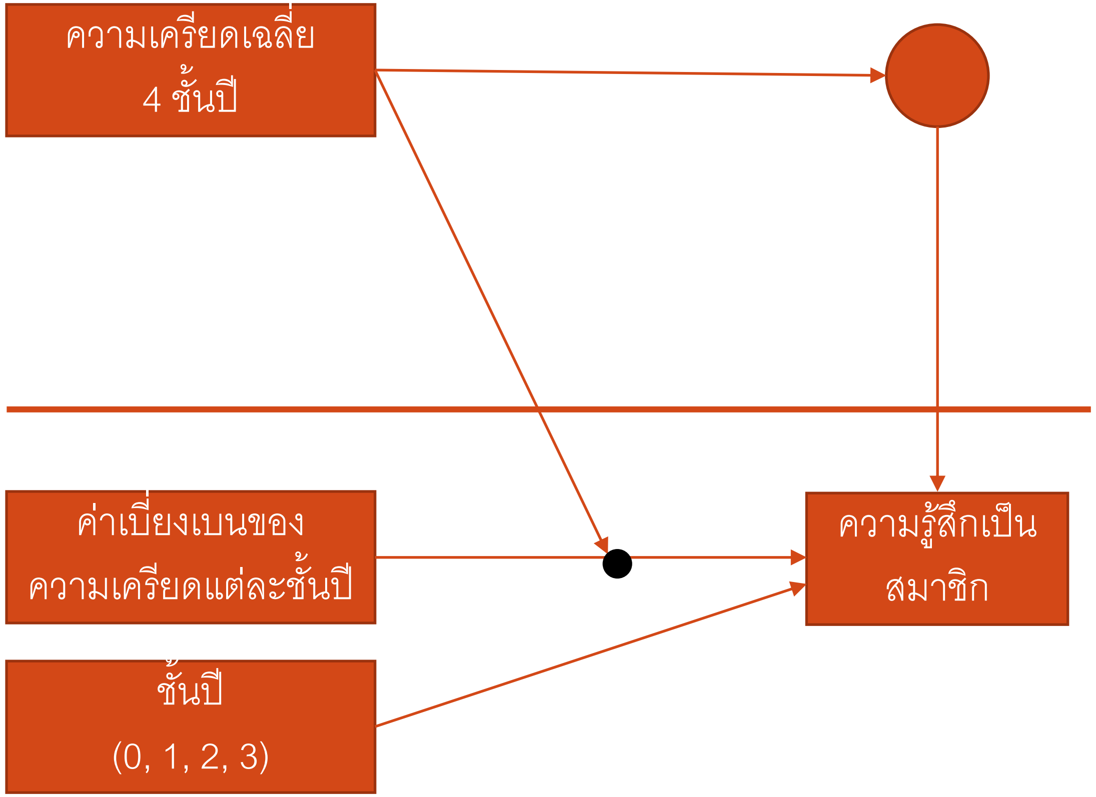
ความสัมพันธ์ในตัว (Autocorrelation)

- การวิเคราะห์ต้องใช้ `nlme` package เพื่อให้สามารถประมาณค่าความสัมพันธ์ในตัว (Autocorrelation) ได้

ความสัมพันธ์ในตัว (Autocorrelation)

ทำนายความรู้สึกเป็นสมาชิก
ของนิสิตจากคณะแห่งหนึ่ง

เก็บข้อมูลจากนิสิต 850 คน เป็นนิสิตรุ่น 2550 ทั้งหมด 426 คน และนิสิตรุ่น 2551 จำนวน 424 คน เก็บข้อมูลระยะยาวตลอดเวลา 4 ปี



```
> out5e <- lme(mem ~ 1 + timec + diffstress + avestress,
+             random = ~1 + timec|pid,
+             data=dat5, method="ML", na.action=na.omit,
+             correlation=corARMA(form = ~ 1 + timec | pid, p = 1))
> summary(out5e)
Linear mixed-effects model fit by maximum likelihood
  Data: dat5
       AIC      BIC    logLik
  25546.58 25601.76 -12764.29

Random effects:
 Formula: ~1 + timec | pid
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 15.040698 (Intr)
timec        5.761743 -0.133
Residual     5.762396

Correlation Structure: AR(1)
 Formula: ~1 + timec | pid
 Parameter estimate(s):
      Phi
0.2107785
Fixed effects:  mem ~ 1 + timec + diffstress + avestress
               Value Std.Error   DF  t-value p-value
(Intercept) 56.73589 2.5786525 2548 22.00214  0.0000
timec       -1.87752 0.2177654 2548 -8.62177  0.0000
diffstress   0.40889 0.0120720 2548 33.87091  0.0000
avestress    0.10181 0.0501331  848  2.03083  0.0426
```

เพิ่มบรรทัดนี้มา เพื่อให้มี AR(1) ระหว่างค่าคงเหลือระดับที่ 1 โดยค่าคงเหลือจากเวลาที่ `timec` ใกล้กันภายในคน (`pid`) เดียวกัน จะมีอิทธิพลต่อกัน

$$e_{tj} = 0.211e_{t-1,j} + r_{tj}$$

$$\sigma_r = 5.76$$

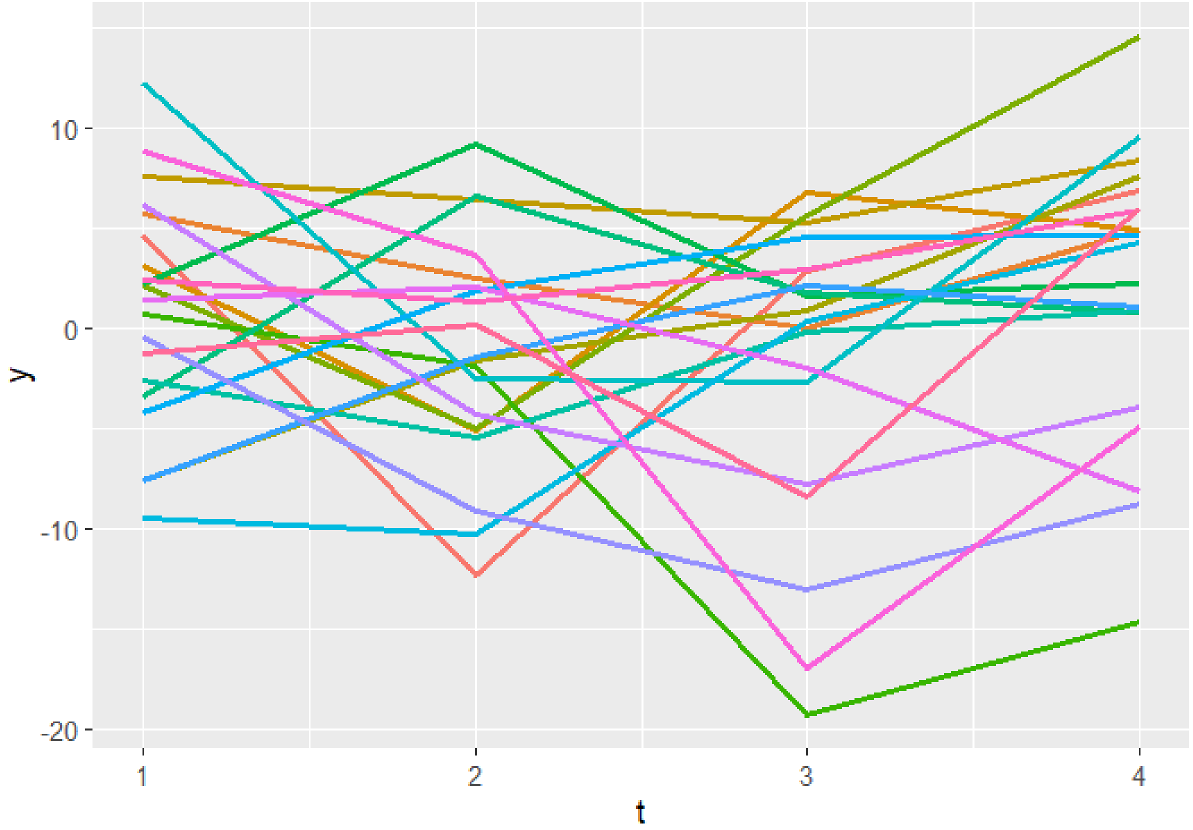
```
> anova(out5a, out5e)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
out5a     1  8 25552.95 25602.00 -12768.47
out5e     2  9 25546.58 25601.76 -12764.29 1 vs 2 8.370098  0.0038
```

เทียบกับโมเดลที่ไม่มี AR(1) พบว่าโมเดลที่มี AR(1) อธิบายข้อมูลได้ดีกว่าอย่างมีนัยสำคัญ

```
> set.seed(123321)
> ar1 <- coef(out5e$modelStruct$corStruct, uncons = FALSE)
> out5esigma <- as.numeric(VarCorr(out5e)["Residual", "StdDev"])
> e1 <- rnorm(20, 0, out5esigma/ sqrt(1 - ar1^2))
> e2 <- ar1*e1 + rnorm(20, 0, out5esigma)
> e3 <- ar1*e2 + rnorm(20, 0, out5esigma)
> e4 <- ar1*e3 + rnorm(20, 0, out5esigma)
>
> mye <- c(e1, e2, e3, e4)
> myid <- rep(1:20, 4)
> myt <- rep(1:4, each=20)
> daterror5 <- data.frame(t=myt, y=mye, id=myid)
> library(ggplot2)
> g <- ggplot(data=daterror5, aes(x=t, y=y, group=id))
> g + geom_line(aes(color=factor(id)), size=1)
```

$$e_{tj} = 0.211 e_{t-1,j} + r_{tj}$$

$$\sigma_r = 5.76$$



```
> out5f <- lme(mem ~ 1 + timec + diffstress + avestress,
+             random = ~1 + timec|pid,
+             data=dat5, method="ML", na.action=na.omit,
+             correlation=corARMA(form = ~ 1 + timec | pid, p = 2))
> summary(out5f)
Linear mixed-effects model fit by maximum likelihood
  Data: dat5
       AIC      BIC    logLik
  25547.67 25608.98 -12763.83

Random effects:
 Formula: ~1 + timec | pid
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 14.017530 (Intr)
timec        5.522065 -0.107
Residual     7.838444

Correlation Structure: ARMA(2,0)
 Formula: ~1 + timec | pid
 Parameter estimate(s):
     Phi1      Phi2
0.4761667 0.1355471
Fixed effects:  mem ~ 1 + timec + diffstress + avestress
               Value Std.Error   DF  t-value p-value
(Intercept) 56.64618 2.5773947 2548 21.97808  0.0000
timec       -1.87812 0.2178019 2548 -8.62307  0.0000
diffstress   0.40870 0.0120714 2548 33.85686  0.0000
avestress    0.10435 0.0501045  848  2.08274  0.0376
```

เพิ่มบรรทัดนี้มา เพื่อให้มี AR(2) ระหว่างค่าคงเหลือ ระดับที่ 1 โดยค่าคงเหลือจากเวลาที่ `timec` ใกล้กัน ภายในคน (`pid`) เดียวกัน จะมีอิทธิพลต่อกัน

$$e_{tj} = 0.476e_{t-1,j} + 0.136e_{t-2,j} + r_{tj}$$

$$\sigma_r = 5.76$$

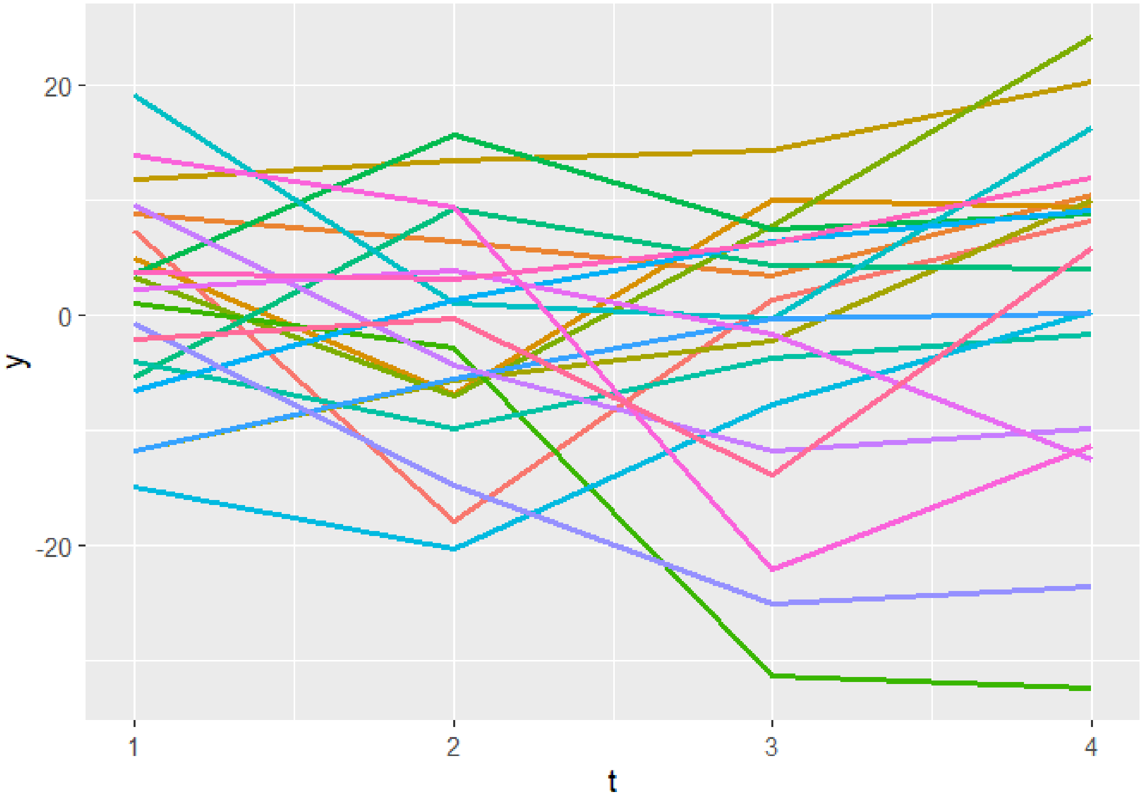
```
> anova(out5e, out5f)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
out5e     1  9 25546.58 25601.76 -12764.29
out5f     2 10 25547.67 25608.99 -12763.83 1 vs 2 0.9078822  0.3407
```

เทียบกับโมเดล AR(1) พบว่า AR(2) อธิบายข้อมูลไม่แตกต่างจาก AR(1) อย่างมีนัยสำคัญ

```
> set.seed(123321)
> ar <- coef(out5f$modelStruct$corStruct, uncons = FALSE)
> out5fsigma <- as.numeric(VarCorr(out5f)["Residual", "StdDev"])
> out5fsigma <- out5fsigma/sqrt(1 - ar[1]^2 - ar[2]^2 - ((ar[1]^2)*ar[2]/(1 - ar[2])))
> e1 <- rnorm(20, 0, out5fsigma)
> e2 <- ar[1]*e1 + rnorm(20, 0, out5fsigma)
> e3 <- ar[1]*e2 + ar[2]*e1 + rnorm(20, 0, out5fsigma)
> e4 <- ar[1]*e3 + ar[2]*e2 + rnorm(20, 0, out5fsigma)
>
> mye <- c(e1, e2, e3, e4)
> myid <- rep(1:20, 4)
> myt <- rep(1:4, each=20)
> daterror6 <- data.frame(t=myt, y=mye, id=myid)
> library(ggplot2)
> g <- ggplot(data=daterror6, aes(x=t, y=y, group=id))
> g + geom_line(aes(color=factor(id)), size=1)
```

$$e_{tj} = 0.476e_{t-1,j} + 0.136e_{t-2,j} + r_{tj}$$

$$\sigma_r = 5.76$$



```
> out5g <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit,
+              weights=varIdent(form = ~1|timec),
+              correlation=corARMA(form = ~ 1 + timec | pid, p = 1))
> summary(out5g)
Linear mixed-effects model fit by maximum likelihood
  Data: dat5
       AIC      BIC    logLik
  25493.77 25567.35 -12734.89

Random effects:
 Formula: ~1 + timec | pid
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 14.711756 (Intr)
timec        5.794574 -0.108
Residual     7.054706

Correlation Structure: AR(1)
 Formula: ~1 + timec | pid
 Parameter estimate(s):
      Phi
0.2803977

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | timec
 Parameter estimates:
        0         1         2         3
1.0000000 0.7982946 0.9267197 0.1845082
Fixed effects:  mem ~ 1 + timec + diffstress + avestress
               Value Std.Error   DF  t-value p-value
(Intercept) 56.74831 2.5775902 2548 22.01603  0.0000
timec       -1.79050 0.2109161 2548 -8.48917  0.0000
diffstress   0.40719 0.0119571 2548 34.05390  0.0000
avestress    0.10043 0.0501128  848  2.00404  0.0454
```

ใส่ทั้ง AR(1) และให้ความแปรปรวนแตกต่างกันระหว่างช่วงเวลา

$$e_{tj} = 0.280e_{t-1,j} + r_{tj}$$

$$\sigma_{r_1} = 7.055$$

$$\sigma_{r_2} = 7.055 \times 0.798 = 5.630$$

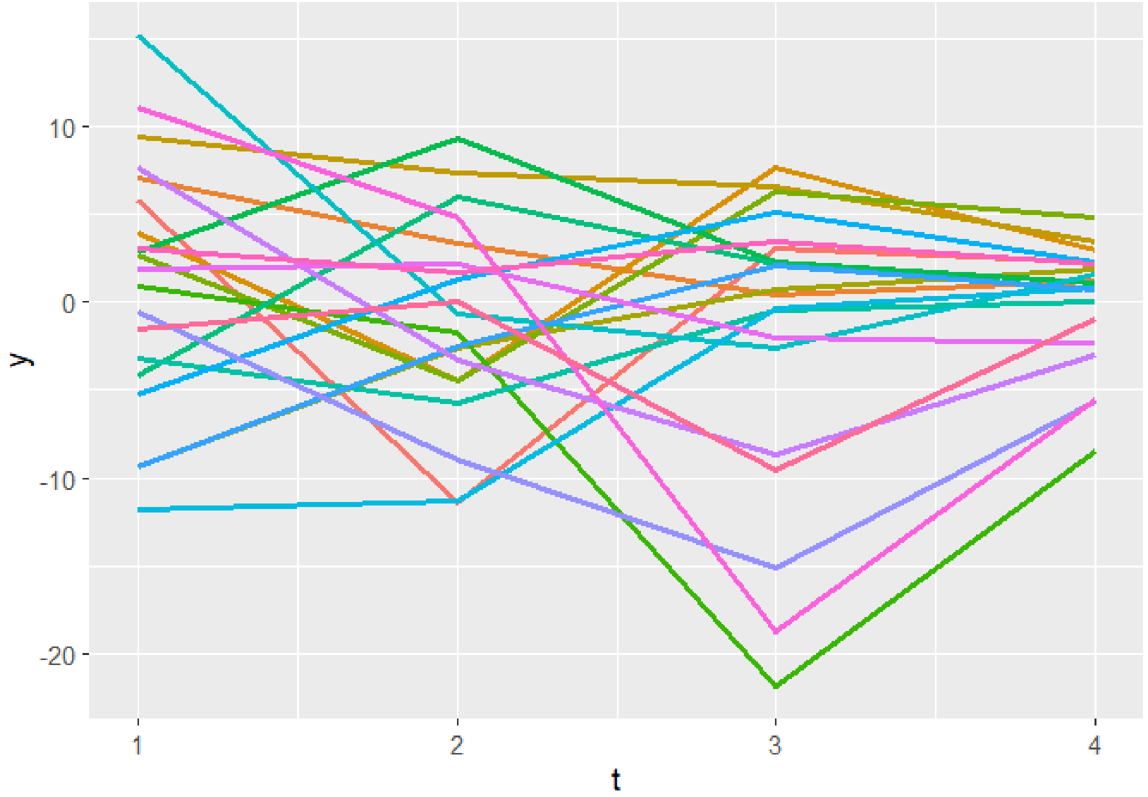
$$\sigma_{r_3} = 7.055 \times 0.927 = 6.540$$

$$\sigma_{r_4} = 7.055 \times 0.185 = 1.305$$

```
> set.seed(123321)
> ar1 <- coef(out5g$modelStruct$corStruct, uncons = FALSE)
> sdratio <- coef(out5g$modelStruct$varStruct, uncons = FALSE)
> out5gsigma <- as.numeric(VarCorr(out5g)["Residual", "StdDev"])
> e1 <- rnorm(20, 0, out5gsigma/sqrt(1 - ar1^2))
> e2 <- ar1*e1 + rnorm(20, 0, out5gsigma * sdratio[1])
> e3 <- ar1*e2 + rnorm(20, 0, out5gsigma * sdratio[2])
> e4 <- ar1*e3 + rnorm(20, 0, out5gsigma * sdratio[3])
>
> mye <- c(e1, e2, e3, e4)
> myid <- rep(1:20, 4)
> myt <- rep(1:4, each=20)
> daterror5 <- data.frame(t=myt, y=mye, id=myid)
> library(ggplot2)
> g <- ggplot(data=daterror5, aes(x=t, y=y, group=id))
> g + geom_line(aes(color=factor(id)), size=1)
```

$$e_{tj} = 0.280e_{t-1,j} + r_{tj}$$

$$\sigma_{r_1} = 7.055$$
$$\sigma_{r_2} = 5.630$$
$$\sigma_{r_3} = 6.540$$
$$\sigma_{r_4} = 1.305$$



สัมประสิทธิ์การทำนาย

- ตัวแปรเวลามักจะไม่ใช้การย้ายศูนย์กลางไปที่ค่าเฉลี่ยกลุ่ม จึงไปใช้วิธีหาสัมประสิทธิ์การทำนายจากโมเดลที่ไม่ได้ย้ายศูนย์กลางไปที่ค่าเฉลี่ยกลุ่มได้
- การย้ายศูนย์กลางของตัวแปรเวลาด้วยค่าคงที่ ไม่มีผลต่อสัมประสิทธิ์ทำนาย
- ยกตัวอย่างการเปลี่ยนแปลงความรู้สึกเป็นสมาชิก ตลอด 4 ชั้นปี
- ใช้ผลการวิเคราะห์จากคาบเรียนที่แล้ว เลือกโมเดลที่ไม่มีความสัมพันธ์ในตัว และความแปรปรวนของค่าคงเหลือเท่ากันในทุกช่วงเวลา

```
> dat5 <-read.table("lecture8ex2.csv", sep=",", header=TRUE)
> dat5$timec <- dat5$time - 1
> dat5$avestress <- ave(dat5$stress, dat5$pid)
> dat5$diffstress <- dat5$stress - dat5$avestress
> library(nlme)
> out5a <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit)
> sumout5a <- summary(out5a)
>
> r2mlm:::r2mlm_long_manual(data=dat5,
+                           covs=c("timec", "diffstress", "avestress"),
+                           random_covs=c("timec"),
+                           gammas=coef(sumout5a)[-1, "Value"],
+                           clusterID="pid",
+                           Tau=getVarCov(out5a),
+                           sigma2=out5a$sigma^2)
$Decompositions
```

แบ่งความแปรปรวนออกเป็นส่วนต่างๆ

	total	within	between
fixed slopes (within)	0.0471088594128234	0.195211606767213	NA
fixed slopes (between)	0.00316696581884948	NA	0.00417432152419909
slope variation (within)	0.122351796082766	0.507006346616612	NA
slope variation (between)	0	NA	0
intercept variation (between)	0.755511013462372	NA	0.995825678475801
residual (within)	0.0718613652231889	0.297782046616175	NA

$R2s

	total	within	between
f1	0.0471088594128234	0.195211606767213	NA
f2	0.00316696581884948	NA	0.00417432152419909
v1	0.122351796082766	0.507006346616612	NA
v2	0	NA	0
m	0.755511013462372	NA	0.995825678475801
f	0.0502758252316729	NA	NA
fv	0.172627621314439	0.702217953383825	0.00417432152419909
fvm	0.928138634776811	NA	NA

ผลของ R^2 ที่ได้

$$R_t^{2(f_1)} = .04$$
$$R_t^{2(f_2)} = .003$$
$$R_t^{2(v_1)} = .122$$
$$R_t^{2(v_2)} = .000$$
$$R_t^{2(m)} = .756$$

ตัวแปรต้นทั้งหมดโดยกำหนดให้ความชันแบบแบบถาวรสามารถอธิบายความแปรปรวนได้ 5%, $R_t^{2(f)} = .05$

สัมประสิทธิ์การทำนาย

- Rights & Sterba (2021) ยังได้พิสูจน์เพิ่มเติมว่า โครงสร้างค่าสหสัมพันธ์ระหว่างค่าคงเหลือ เช่น ความสัมพันธ์ในตัว ไม่มีผลต่อสูตรการหาสัมประสิทธิ์การทำนาย
  - อย่างไรก็ตาม ผู้วิเคราะห์ต้องใส่ $Var(e_{tj})$ ไม่ใช่ $Var(r_{tj})$ ซึ่งในโมเดล AR(1) จะมีค่าดังนี้

$$Var(e_{tj}) = \frac{Var(r_{tj})}{1-\rho^2}$$

- ใช้ตัวอย่างเดิม แต่ปรับให้มี AR(1) และความแปรปรวนของค่าคงเหลือเท่ากันในทุกช่วงเวลา

```
> out5e <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit,
+              correlation=corARMA(form = ~ 1 + timec | pid, p = 1))
> sumout5e <- summary(out5e)
>
> ar1 <- coef(out5e$modelStruct$corStruct, uncons = FALSE)
>
> r2mlm::r2mlm_long_manual(data=dat5,
+                          covs=c("timec", "diffstress", "avestress"),
+                          random_covs=c("timec"),
+                          gammas=coef(sumout5e)[-1, "Value"],
+                          clusterID="pid",
+                          Tau=getVarCov(out5e),
+                          sigma2=out5e$sigma^2/(1 - ar1^2))
$Decompositions
```

หาสัมประสิทธิ์ถดถอยในตัว AR(1)

ใช้สูตรเพื่อใส่ $Var(e_{ij})$

	total	within	between
fixed slopes (within)	0.0470165037236248	0.181952950274654	NA
fixed slopes (between)	0.00324887388858962	NA	0.00438089340729464
slope variation (within)	0.11506090083776	0.445283436891759	NA
slope variation (between)	0	NA	0
intercept variation (between)	0.738351887997082	NA	0.995619106592705
residual (within)	0.0963218335529436	0.372763612833588	NA

แบ่งความแปรปรวนออกเป็นส่วนต่างๆ

$R2s

	total	within	between
f1	0.0470165037236248	0.181952950274654	NA
f2	0.00324887388858962	NA	0.00438089340729464
v1	0.11506090083776	0.445283436891759	NA
v2	0	NA	0
m	0.738351887997082	NA	0.995619106592705
f	0.0502653776122144	NA	NA
fv	0.165326278449975	0.627236387166412	0.00438089340729464
fvm	0.903678166447056	NA	NA

ผลของ R^2 ที่ได้

$$R_t^{2(f_1)} = .047$$
$$R_t^{2(f_2)} = .003$$
$$R_t^{2(v_1)} = .115$$
$$R_t^{2(v_2)} = .000$$
$$R_t^{2(m)} = .738$$

ตัวแปรต้นทั้งหมดโดยกำหนดให้ความชันแบบแบบถาวรสามารถอธิบายความแปรปรวนได้ 5%, $R_t^{2(f)} = .047$

สัมประสิทธิ์การทำนาย

- โมเดลค่าความแปรปรวนแตกต่างกัน (Heteroscedastic Model) มีผลต่อการคำนวณสัมประสิทธิ์ทำนาย เพราะมีผลต่อการหาสัดส่วนความแปรปรวนคงเหลือระดับที่ 1
- Rights & Sterba (2021) แสดงให้เห็นว่า ผู้วิเคราะห์ต้องหาค่าเฉลี่ยความแปรปรวนของทั้งกลุ่มตัวอย่างหรือ $E\left(\sigma_i^2\right)$ เพื่อใส่เข้าไปในสูตรการคำนวณสัมประสิทธิ์การทำนาย ซึ่งค่อนข้างซับซ้อน

สัมประสิทธิ์การทำนาย

- คำแนะนำของผม คือ คำนวณค่าคาดหวังของความแปรปรวนจากทุกๆ หน่วยของตัวอย่างที่ 1 แล้วมาหารเฉลี่ยจากทั้งกลุ่มตัวอย่าง จะเป็นวิธีประมาณค่าที่ง่ายที่สุด
- ตัวอย่างโมเดลที่ 3 โมเดลที่ไม่มีความสัมพันธ์ในตัว และความแปรปรวนแตกต่างกันในแต่ละช่วงเวลา

```
> out5b <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit,
+              weights=varIdent(form = ~1|timec))
> sumout5b <- summary(out5b)
>
> out5bsigma <- out5b$sigma
> out5bsigmatime <- out5bsigma * coef(out5b$modelStruct$varStruct, uncons = FALSE)
> out5berrvar <- c(out5bsigma, out5bsigmatime)^2
> out5berrvar <- cbind(1:4, out5berrvar)
>
> matchindex5 <- match(dat5$time, out5berrvar[,1])
> dat5$vare <- out5berrvar[matchindex5, 2]
> aveerrvar5b <- mean(dat5$vare)
```

หาความแปรปรวนของค่าคงเหลือทุกช่วงเวลา

สร้างเป็นตาราง

เอาตารางความแปรปรวนของแต่ละช่วงเวลา ไปจับคู่กับเวลาในแต่ละแถว แล้วนำความแปรปรวนมาใส่ในคอลัมน์ใหม่

หาค่าเฉลี่ยจากทุกแถว

```
> r2mlm:::r2mlm_long_manual(data=dat5,
+                           covs=c("timec", "diffstress", "avestress"),
+                           random_covs=c("timec"),
+                           gammas=coef(sumout5b)[-1, "Value"],
+                           clusterID="pid",
+                           Tau=getVarCov(out5b),
+                           sigma2=aveerrvar5b)
$Decompositions
                             total                within            between
fixed slopes (within)        0.0458638418328702   0.192808967631146 NA
fixed slopes (between)       0.00290159730400247  NA                0.00380723060820706
slope variation (within)     0.121262016573532    0.509778581428324 NA
slope variation (between)    0                    NA                0
intercept variation (between) 0.759226469681916   NA                0.996192769391793
residual (within)            0.0707460746076797   0.29741245094053  NA
```

แบ่งความแปรปรวนออกเป็นส่วนต่างๆ

```
$R2s
    total               within            between
f1  0.0458638418328702  0.192808967631146 NA
f2  0.00290159730400247 NA                0.00380723060820706
v1  0.121262016573532   0.509778581428324 NA
v2  0                   NA                0
m   0.759226469681916   NA                0.996192769391793
f   0.0487654391368726  NA                NA
fv  0.170027455710405   0.70258754905947  0.00380723060820706
fvm 0.92925392539232    NA                NA
```

ผลของ R^2 ที่ได้

$$R_t^{2(f_1)} = .046$$
$$R_t^{2(f_2)} = .003$$
$$R_t^{2(v_1)} = .121$$
$$R_t^{2(v_2)} = .000$$
$$R_t^{2(m)} = .759$$

ตัวแปรต้นทั้งหมดโดยกำหนดให้ความชันแบบแบบถาวรสามารถอธิบายความแปรปรวนได้ 5%, $R_t^{2(f)} = .046$

สัมประสิทธิ์การทำนาย

- สำหรับโมเดลที่มีทั้ง AR(1) และความแปรปรวนแตกต่างกันในแต่ละช่วงเวลา จะซับซ้อน เนื่องจากเราไม่สามารถเหมาได้ว่า ความแปรปรวน ณ เวลาก่อนหน้าเท่ากับความแปรปรวนเวลาปัจจุบัน

$$Var(e_{tj}) = \rho^2 Var(e_{t-1,j}) + Var(r_{tj})$$

- ดังนั้น เราไม่สามารถหา $Var(e_{tj})$ ณ จุดเริ่มต้นได้เลย เพราะเราไม่ทราบ $Var(e_{t-1,j})$ หรือความแปรปรวนค่าคงเหลือ ณ เวลาก่อนหน้า เนื่องจากไม่ได้วัด
  - ดังนั้น สิ่งที่พอทำได้ จึงเป็นการเหมาว่า เวลาก่อนจุดเริ่มต้น มีความแปรปรวนเท่ากับจุดเริ่มต้น

สัมประสิทธิ์การทำนาย

- จากการเหมาแบบนี้ ทำให้

$$Var(e_1\ ) = \frac{\sigma_1^2}{1-\rho^2}$$

- ส่งผลให้

$$Var(e_2) = \rho^2 Var(e_1) + \sigma_2^2$$
$$Var(e_3) = \rho^2 Var(e_2) + \sigma_3^2$$
$$Var(e_t) = \rho^2 Var(e_{t-1}) + \sigma_t^2$$

- เมื่อได้แบบนี้แล้ว ให้นำค่าคาดหวังไปใส่ในทุกหน่วยของกลุ่มตัวอย่าง เพื่อหาค่าเฉลี่ย หา $E\left(\sigma^2\right)$

สัมประสิทธิ์การทำนาย

- ตัวอย่างโมเดลที่ 4 โมเดล AR(1) และความแปรปรวนแตกต่างกันในแต่ละช่วงเวลา

```
> out5g <- lme(mem ~ 1 + timec + diffstress + avestress,
+              random = ~1 + timec|pid,
+              data=dat5, method="ML", na.action=na.omit,
+              weights=varIdent(form = ~1|timec),
+              correlation=corARMA(form = ~ 1 + timec | pid, p = 1))
> sumout5g <- summary(out5g)
>
> ar1 <- coef(out5g$modelStruct$corStruct, uncons = FALSE)
> sdratio <- coef(out5g$modelStruct$varStruct, uncons = FALSE)
> out5gsigmatime <- out5g$sigma * sdratio
> out5gerrvar <- c(out5gsigma, out5gsigmatime)^2
>
> errvar1g <- out5gerrvar[1] / (1 - ar1^2)
> errvar2g <- (ar1^2)*errvar1g + out5gerrvar[2]
> errvar3g <- (ar1^2)*errvar2g + out5gerrvar[3]
> errvar4g <- (ar1^2)*errvar3g + out5gerrvar[4]
>
> out5gerrvar <- cbind(1:4, c(errvar1g, errvar2g, errvar3g, errvar4g))
>
> matchindex5 <- match(dat5$time, out5gerrvar[,1])
> dat5$varg <- out5gerrvar[matchindex5, 2]
> aveerrvar5g <- mean(dat5$varg)
```

หา $Var(r_t)$ ทุกช่วงเวลา

หา $Var(e_1)$

หา $Var(e_2), Var(e_3), Var(e_4)$

สร้างเป็นตาราง

เอาตารางความแปรปรวนของแต่ละช่วงเวลาไปจับคู่กับเวลาในแต่ละแถว แล้วนำความแปรปรวนมาใส่ในคอลัมน์ใหม่

หาค่าเฉลี่ยจากทุกแถว

```
> r2mlm::r2mlm_long_manual(data=dat5,
+                          covs=c("timec", "diffstress", "avestress"),
+                          random_covs=c("timec"),
+                          gammas=coef(sumout5g)[-1, "Value"],
+                          clusterID="pid",
+                          Tau=getVarCov(out5g),
+                          sigma2=averrvar5g)
$Decompositions
                              total               within            between
fixed slopes (within)         0.0457801518487724  0.175599608528289 NA
fixed slopes (between)        0.00317538573831712 NA                0.00429516889715021
slope variation (within)      0.116899028145694   0.448391338838945 NA
slope variation (between)     0                   NA                0
intercept variation (between) 0.736117017972268   NA                0.99570483110285
residual (within)             0.0980284162949485  0.376009052632766 NA
```

แบ่งความแปรปรวนออกเป็นส่วนต่างๆ

```
$R2s
    total               within            between
f1  0.0457801518487724  0.175599608528289 NA
f2  0.00317538573831712 NA                0.00429516889715021
v1  0.116899028145694   0.448391338838945 NA
v2  0                   NA                0
m   0.736117017972268   NA                0.99570483110285
f   0.0489555375870895  NA                NA
fv  0.165854565732783   0.623990947367234 0.00429516889715021
fvm 0.901971583705051   NA                NA
```

ผลของ R^2 ที่ได้

$R_t^{2(f_1)} = .046$

$R_t^{2(f_2)} = .003$

$R_t^{2(v_1)} = .117$

$R_t^{2(v_2)} = .000$

$R_t^{2(m)} = .736$

ตัวแปรต้นทั้งหมดโดยกำหนดให้ความชันแบบแบบถาวรสามารถอธิบายความแปรปรวนได้ 5%, $R_t^{2(f)} = .046$

สัมประสิทธิ์การทำนาย

- จากตัวอย่างโมเดลที่ 1 ถึง 4 ตัวแปรความเครียดและตัวแปรเวลา สามารถอธิบายความแปรปรวนได้ประมาณ 5% เมื่อให้ขนาดอิทธิพลคงที่
- หากดูภายในแต่ละบุคคล ค่าเบี่ยงเบนความเครียดและตัวแปรเวลา สามารถอธิบายความแปรปรวนของการเปลี่ยนแปลงภายในบุคคลได้ $R_w^{2(fv)} = 62\%$
- ความแปรปรวนส่วนใหญ่ เป็นความแตกต่างระหว่างบุคคล (m)
- แม้โมเดลที่ 1 ถึง 4 ดูแล้วค่าไม่ค่อยแตกต่างกัน แต่อาจเป็นเฉพาะตัวอย่างนี้ ตัวอย่างอื่นอาจมีค่าแตกต่างกัน

คาบต่อไป

- โมเดล 3 ระดับและโมเดลข้อมูลเชิงซ้อนแบบอื่น
- การบ้านที่ 9