



คู่มือการใช้ R ในการวิเคราะห์สถิติสำหรับการทดลอง

โดย

ผศ. ดร. ปิยะ กิตติภาดากุล

ภาควิชาพืชไร่นา คณะเกษตร มหาวิทยาลัยเกษตรศาสตร์

สำหรับ

การฝึกอบรม การพัฒนาศักยภาพบุคลากร รุ่นที่ 4

หลักสูตร เทคนิคการวิจัยด้านป้าไม้

วันที่ 13 กุมภาพันธ์ 2561

กรมป่าไม้

สารบัญ

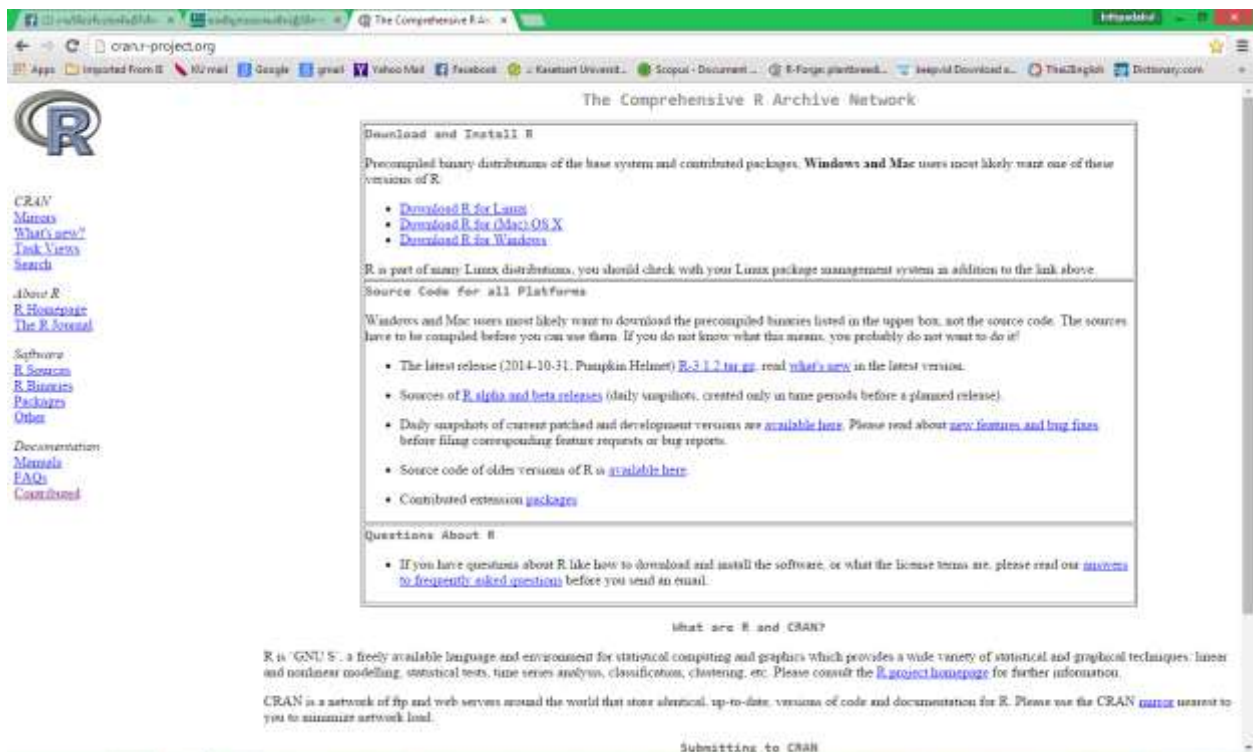
	หน้า
คำนำ	i
กำหนดการฝึกอบรม	ii
สารบัญ	iii
โปรแกรม R	1
การติดตั้งโปรแกรม R	1
การวิเคราะห์ความแปรปรวน	4
แผนการทดลองแบบสุ่มตลอด (Completely Randomized Design; CRD)	5
แผนการทดลองแบบสุ่มในบล็อกสมบูรณ์ (Randomized Complete Block Design; RCBD)	8
การใช้ package อื่นมาวิเคราะห์ร่วมกับ R	11
การบันทึก Object ของผลการวิเคราะห์จาก R	14
การทดลองแบบแฟคตอเรียลในแผนการทดลองแบบสุ่มตลอด (Factorial in CRD)	14
บรรณานุกรม	18
สไลด์บรรยาย การวางแผนการทดลอง	19
สไลด์บรรยาย แนะนำ R	28

โปรแกรม R

โปรแกรม R เริ่มแรกพัฒนาโดย Ross Ihaka และ Robert Gentleman จาก University of Auckland ประเทศนิวซีแลนด์ R เป็นโปรแกรมที่ใช้โดยไม่เสียค่าใช้จ่าย สามารถโหลดจากอินเทอร์เน็ต และใช้ในคอมพิวเตอร์ของผู้ใช้โดยไม่จำเป็นต้องมีการเชื่อมต่อกับอินเทอร์เน็ตหลังการติดตั้ง และยังเปิดโอกาสให้ผู้ใช้สามารถปรับปรุงพัฒนาให้โปรแกรมมีประสิทธิภาพยิ่งขึ้น (open source) จึงเป็นโปรแกรมที่นิยมใช้โดยนักสถิติมืออาชีพในประเทศต่าง ๆ โดยเฉพาะประเทศในยุโรปและอเมริกา R เป็นโปรแกรมที่มีสมรรถนะสูงในการวิเคราะห์ทางสถิติและการสร้างกราฟในรูปแบบต่าง ๆ ในความเป็นจริง R ไม่ใช่โปรแกรมคอมพิวเตอร์เพียงอย่างเดียว ซึ่ง R เป็นภาษาทางคอมพิวเตอร์ด้วย จึงมีโปรแกรมเมอร์เขียนโปรแกรม (package) ออกมาใช้งานและวิเคราะห์ในสาขาต่าง ๆ ให้ได้เลือกใช่มากมายตามความนิยมหรือถนัดในสาขาเฉพาะ

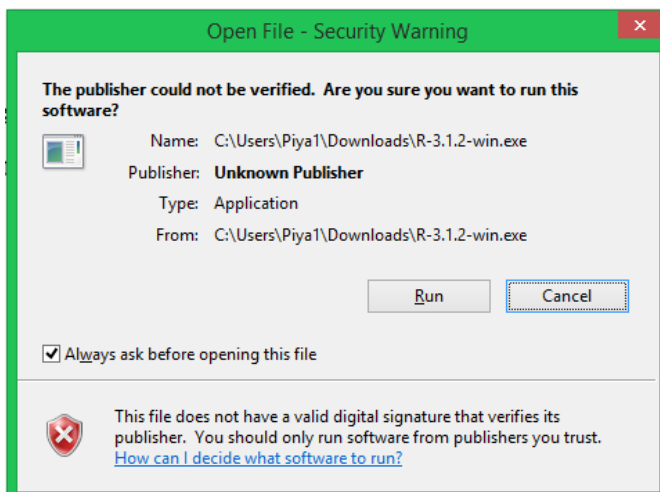
การติดตั้งโปรแกรม R

การติดตั้งโปรแกรม R ให้ไปที่เว็บไซต์ <http://cran.r-project.org/> โดยให้เลือกดาวโหลดตามปฏิบัติการที่ใช้ เมื่ออยู่ที่เว็บไซต์ของ R แล้ว ให้เลือกระบบปฏิบัติการที่ใช้ในคอมพิวเตอร์ของผู้อ่าน ซึ่งการติดตั้ง R สามารถเลือกติดตั้งได้ 3 ระบบปฏิบัติการคือ Linux Mac หรือ Windows เมื่อเลือกระบบปฏิบัติการที่จะติดตั้งแล้ว ให้คลิก [install R for the first time](#) เพื่อดาวโหลดโปรแกรมในการติดตั้งโปรแกรมในคอมพิวเตอร์

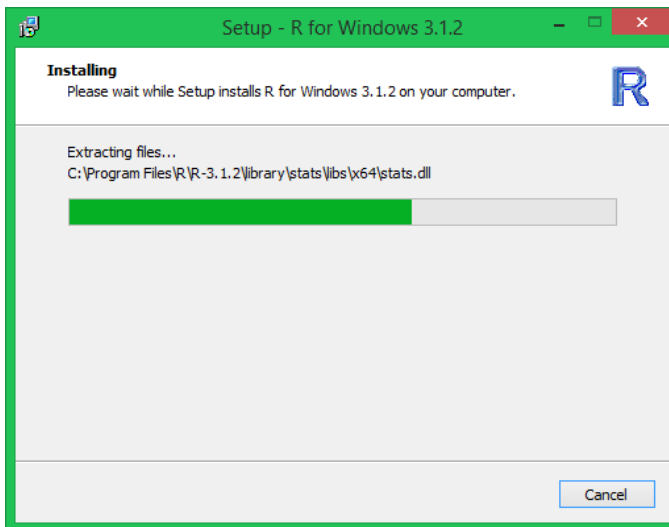




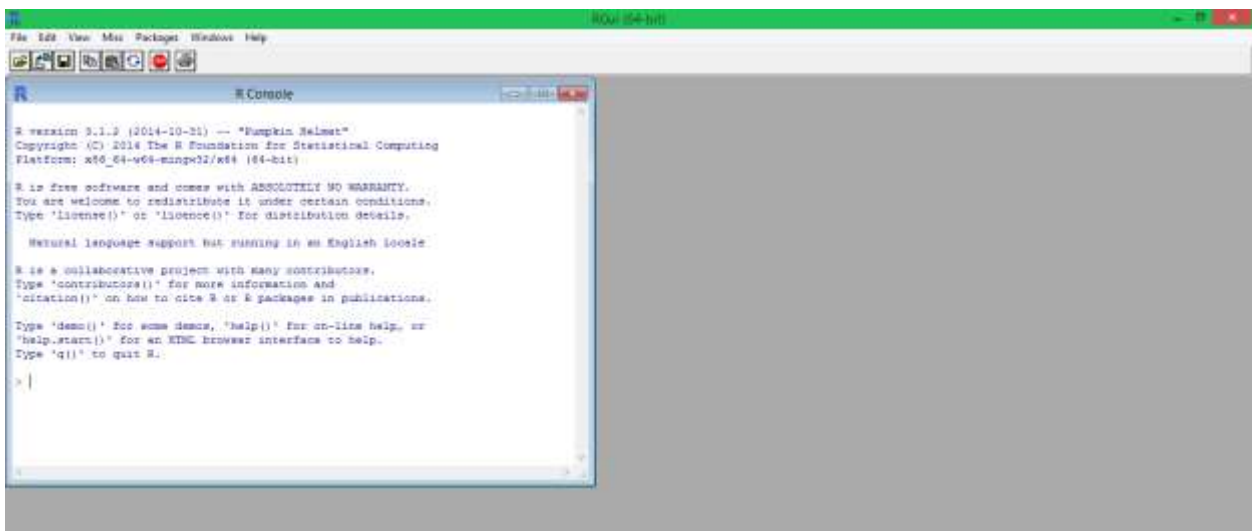
เมื่อคลิกลิงค์ [install R for the first time](#) จะพบลิงค์สำหรับดาวน์โหลด R ในเวอร์ชันล่าสุด ให้คลิกลิงค์ [Download R](#) ในเวอร์ชันล่าสุดที่มีให้โหลด



เมื่อโหลดโปรแกรมแล้ว ให้เลือก Run เพื่อติดตั้ง



เมื่อติดตั้งโปรแกรม R แล้ว การเปิดใช้โปรแกรมจะปรากฏดังภาพข้างล่าง R เป็นโปรแกรมที่ตอบโต้กับผู้ใช้ด้วยการเขียนคำสั่งที่เรียกว่า command-line interface ซึ่งสามารถใช้คำนวณตัวเลขอย่างง่ายจนถึงการวิเคราะห์ขั้นสูง หากต้องการให้จอภาพว่างโดยไม่มีข้อความที่แสดงในตอนเข้าใช้โปรแกรม ให้เลือก Control + L



ซึ่งขณะนี้ผู้อ่านมี R พร้อมสำหรับคำนวณหรือวิเคราะห์ผลทางสถิติต่าง ๆ แล้ว

การวิเคราะห์ความแปรปรวน

การวิเคราะห์ความแปรปรวน (Analysis of Variance; ANOVA) เป็นมาตรการทางเลขาคณิต สำหรับแบ่งแยกความแปรปรวนหรือความแตกต่างระหว่างค่าสังเกตในข้อมูลออกเป็น ส่วน ๆ ตามแหล่งที่มา หรือตามปัจจัยที่ก่อให้เกิดความแปรปรวนนั้น เพื่อตัดสินว่าความแปรปรวนในคุณลักษณะขนาดเท่าไรที่อธิบายด้วยปัจจัยใดบ้าง ประโยชน์ของการวิเคราะห์ความแปรปรวน คือ สามารถทดสอบสมมติฐานว่า “ค่าเฉลี่ยเลขคณิตที่มาจากประชากรหลาย ๆ กลุ่มจะมีค่าเท่ากันหรือไม่” นั่นคือการวิเคราะห์ความแปรปรวนใช้สำหรับกรณีที่การทดลองมีทรีทเมนต์มากกว่า 2 ขึ้นไป การวิเคราะห์ทางสถิติในเรื่องการวิเคราะห์ความแปรปรวน สิ่งสำคัญจะต้องทราบถึงรูปแบบทางคณิตศาสตร์หรือโมเดล (model) ของแผนการทดลองที่ใช้ ถึงจะสามารถเขียนรูปแบบการวิเคราะห์ให้ถูกต้องตามแผนการทดลองที่ใช้ได้ ซึ่งแผนการทดลองหลักที่ใช้ในการทดลองได้แก่ CRD RCBD และ LSD

รูปแบบทางสถิติ (statistical model) ของแผนการทดลองที่ควรรู้

แผนการทดลองแบบ Completely Randomized Design (CRD) มีรูปแบบทางสถิติคือ

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

แผนการทดลองแบบ Randomized Complete Block Design (RCBD) มีรูปแบบทางสถิติคือ

$$Y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}$$

แผนการทดลองแบบ Latin Square (LS) มีรูปแบบทางสถิติคือ

$$Y_{ij} = \mu + R_i + C_j + \tau_k + \epsilon_{ijk}$$

แผนการทดลองแบบ Factorial in CRD มีรูปแบบทางสถิติคือ

$$Y_{ijk} = \mu + A_j + B_k + AB_{jk} + \epsilon_{ijk}$$

แผนการทดลองแบบ Factorial in RCBD มีรูปแบบทางสถิติคือ

$$Y_{ijk} = \mu + \beta_i + A_j + B_k + AB_{jk} + \epsilon_{ijk}$$

แผนการทดลองแบบ Split-plot in CRD มีรูปแบบทางสถิติคือ

$$Y_{ijk} = \mu + M_j + (\beta M)_{ij} + S_k + (MS)_{jk} + \epsilon_{ijk}$$

แผนการทดลองแบบ Split-plot in RCBD มีรูปแบบทางสถิติคือ

$$Y_{ijk} = \mu + \beta_i + M_j + (\beta M)_{ij} + S_k + (MS)_{jk} + \epsilon_{ijk}$$

โดยที่

μ เป็นค่าเฉลี่ยรวม

β_i เป็นอิทธิพลของ blocks หรือ replications ของการทดลอง

τ_j , τ_k เป็นอิทธิพลของ Treatment ของการทดลอง

R_i และ C_j เป็นอิทธิพลของ Row และ Column ตามลำดับ

A_j และ B_k เป็นอิทธิพลของปัจจัย A และ B ตามลำดับ

AB_{jk} เป็นอิทธิพลร่วมของปัจจัย A และ B

M_j และ S_k เป็นอิทธิพลของ Main plots และ Subplots ตามลำดับ

$(\beta M)_{ij}$ เป็นความแปรปรวนของ Replication และ Main plots

$(MS)_{jk}$ เป็นอิทธิพลร่วมของ Main plots กับ Subplots

ϵ_{ij} เป็นความคลาดเคลื่อนของการทดลอง

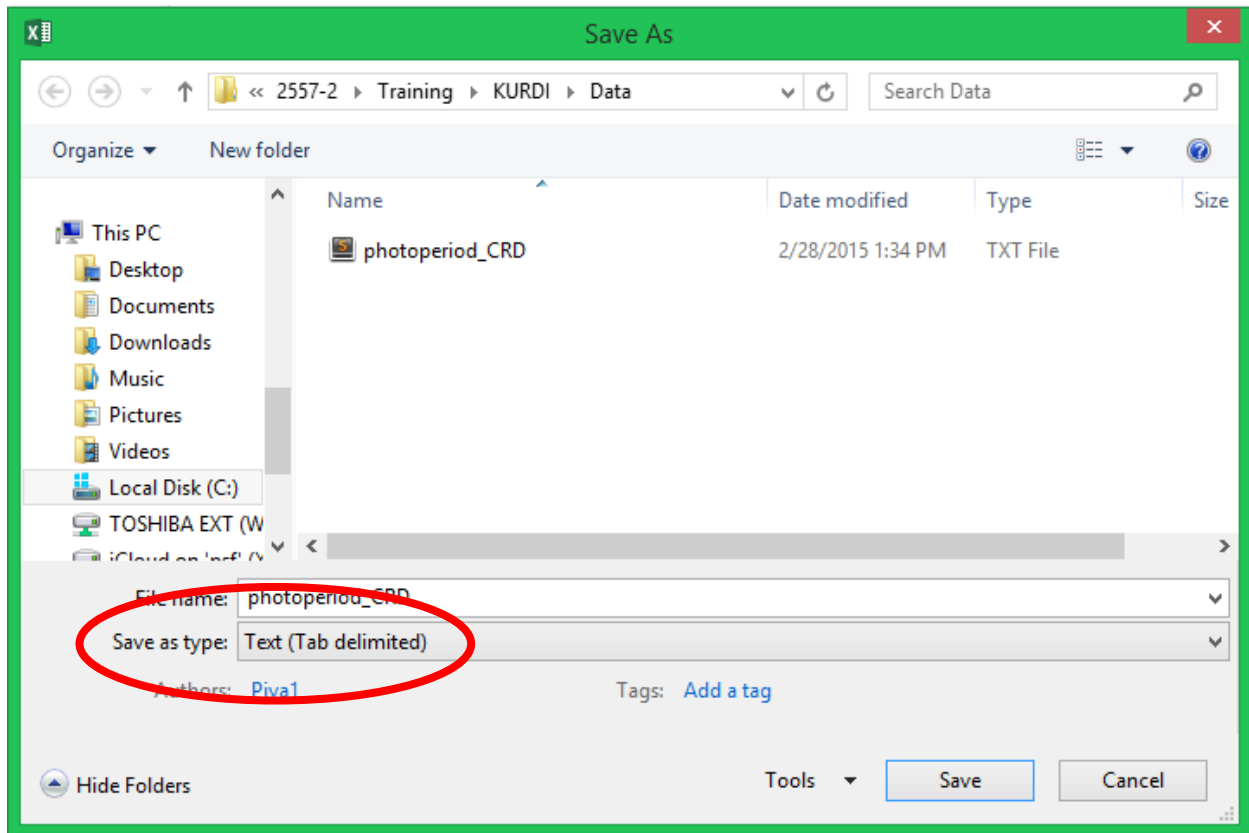
แผนการทดลองแบบสุ่มตลอด (Completely Randomized Design; CRD)

แผนการทดลองแบบสุ่มตลอดเป็นแผนการทดลองที่มีความสะดวกในการปฏิบัติและง่ายสำหรับการวิเคราะห์ สิ่งสำคัญที่ต้องคำนึงถึงในการเลือกใช้แผนการทดลองนี้คือ หน่วยทดลองที่ใช้ต้องมีความสม่ำเสมอ ถ้าการทดลองเลี้ยงปลาเป็นบ่อทดลอง นั่นคือบ่อหรือกระชังที่ใช้เลี้ยงต้องมีความสม่ำเสมอ ไม่มีผลเนื่องมาจากสภาพแวดล้อมต่าง ๆ ที่จะส่งผลต่อทรีทเมนต์ที่ใช้ในการทดลอง นั่นคือไม่มีความแปรปรวนเนื่องจากสภาพแวดล้อมที่จะมีผลต่อทรีทเมนต์ ซึ่งหากมีความแปรปรวนเนื่องจากสภาพแวดล้อมเกิดขึ้นในการทดลอง จะทำให้ผลการทดลองมีความคลาดเคลื่อน หากเกิดกรณีที่มีความแปรปรวนเนื่องจากสภาพแวดล้อมขึ้น วิธีแก้ปัญหาคือต้องเลือกแผนการทดลองอื่นที่สามารถลดหรือจำกัดความแปรปรวนเนื่องจากสภาพแวดล้อมลงเพื่อไม่ให้มีผลกระทบต่อทรีทเมนต์ของการทดลอง และสิ่งสำคัญอีกประการหนึ่งคือทรีทเมนต์ของแผนการทดลองแบบ CRD จะมีเพียงปัจจัยเดียวเท่านั้น

ตัวอย่างที่ 1 การทดลองในตู้เพาะเลี้ยงโดยใช้ช่วงแสง (photoperiod) ที่ต่างกัน 4 ระดับคือ ช่วงแสงสั้นมาก (Very short) ช่วงแสงสั้น (Short) ช่วงแสงยาว (Long) และ ช่วงแสงยาวมาก (Very long) สำหรับการให้แสงแก่ต้นพืชในแต่ละวัน แล้ววัดการตอบสนองต่อช่วงแสงของพืชในด้านการเจริญเติบโต โดยวัดเป็นความสูงที่เพิ่มขึ้นเป็นมิลลิเมตร (mm)

v.short	short	long	v.long
2	3	3	4
3	4	5	6
1	2	1	2
1	1	2	2
2	2	2	2
1	1	2	3

การเตรียมข้อมูลสำหรับการวิเคราะห์สำหรับแผนการทดลองต่าง ๆ ถือเป็นขั้นตอนที่สำคัญในอันดับต้น ๆ เพราะการจัดเรียงข้อมูลโดยแบ่งออกเป็นปัจจัยที่สอดคล้องกับแผนการทดลอง เป็นสิ่งสำคัญต่อความถูกต้องของการนำข้อมูลมาวิเคราะห์ วิธีที่สะดวกสำหรับการใช้โปรแกรมวิเคราะห์ทางสถิติคือ การสร้าง text file (.txt) ใน MS Excel โดยเลือกบันทึกไฟล์เป็นแบบ Text (Tab delimited)



การอ่านข้อมูลจาก text file (.txt)

```
photo = read.table(file.choose(), header=T)
```

การแสดงข้อมูลมีอ่านเข้ามาในโปรแกรม R โดยพิมพ์ชื่อของ object ที่ใช้เก็บข้อมูลสำหรับการวิเคราะห์

photo

```
photoperiod growth
1      v.short      2
2      v.short      3
3      v.short      1
4      v.short      1
5      v.short      2
6      v.short      1
7      short        3
8      short        4
9      short        2
10     short        1
11     short        2
12     short        1
13     long         3
14     long         5
15     long         1
16     long         2
17     long         2
```

```

18      long      2
19     v.long     4
20     v.long     6
21     v.long     2
22     v.long     2
23     v.long     2
24     v.long     3

```

การแนบข้อมูลในโปรแกรม R เพื่อลดขั้นตอนในการอ้างถึงตัวแปรต่าง ๆ

```
attach(photo)
```

การดูชื่อตัวแปรหรือ object ในข้อมูล photo

```
names(photo)
```

```
[1] "photoperiod" "growth"
```

การค่าเฉลี่ยของทรีทเมนต์ในข้อมูล photo

```
tapply(growth, photoperiod, mean)
```

```

      long      short   v.long v.short
2.500000  2.166667  3.166667  1.666667

```

การเรียงค่าเฉลี่ยของทรีทเมนต์ใหม่

```
photoperiod<-ordered(photoperiod, levels=c("v.short", "short", "long", "v.long"))
```

```
tapply(growth, photoperiod, mean)
```

```

  v.short      short      long   v.long
1.666667  2.166667  2.500000  3.166667

```

ก่อนจะวิเคราะห์ความแปรปรวนต้องทราบถึงรูปแบบทางสถิติหรือโมเดลของแผนการทดลอง CRD คือ

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

ในการเขียนโมเดลสำหรับการวิเคราะห์ในโปรแกรมสถิติ มักจะไม่รวมเทอมของ μ และ ε_{ij} ในการวิเคราะห์

ในกรณีนี้ Y_{ij} เป็นความยาวของปลาที่วัดจากการทดลอง

τ_i เป็นสูตรอาหารซึ่งเป็นทรีทเมนต์ของการทดลองนี้

ฉะนั้นกรณีนี้ จึงมี 2 ตัวแปรที่จะเขียนในโมเดลในการวิเคราะห์คือ **photoperiod** และ **growth**

ในที่นี้โมเดลในการวิเคราะห์ตามรูปแบบ Response.variable ~ Explanatory.variable คือ

growth ~ photoperiod

การวิเคราะห์ความแปรปรวน โดยการระบุตัวแปรที่จะวิเคราะห์ตามแผนการทดลอง CRD

```
fit = lm(growth~photoperiod)
```

การแสดงผลของการวิเคราะห์ความแปรปรวน

```
anova(fit)
```

```
Analysis of Variance Table
```

```
Response: growth
```

```

      Df Sum Sq Mean Sq F value Pr(>F)
photoperiod  3   7.125    2.375   1.4615 0.2551
Residuals  20  32.500    1.625

```

ความแปรปรวนของการทดลอง

```
anova(fit) [2,3]
```

```
[1] 1.625
```

การหา critical values of Fisher's F

```
qf(0.95,3,20)
```

```
[1] 3.098391
```

ผลการวิเคราะห์พบว่า $Pr(>F)$ มีค่า > 0.05 (0.2551) นั่นคือ อิทธิพลของช่วงแสงไม่มีผลต่อการเจริญเติบโตของต้นพืช หรือความหมายหนึ่งคือ ช่วงแสงที่แตกต่างกันไม่มีผลให้การเจริญเติบโตของต้นพืชต่างกัน ที่เป็นเช่นนี้เพราะจากค่าของ SSE ซึ่งเป็นความแปรปรวนเนื่องจากการทดลองที่ไม่สามารถอธิบายด้วยปัจจัยจากการทดลองมีค่าสูงมาก (32.5) จึงทำให้อิทธิพลของทรีทเมนต์ (ในที่นี้คือช่วงแสง) ไม่มีความสำคัญ (significant)

เมื่อเสร็จจากการวิเคราะห์ข้อมูล photo ก็ให้ปลดข้อมูลชุดที่กำลังวิเคราะห์

```
detach(photo)
```

แผนการทดลองแบบสุ่มในบล็อกสมบูรณ์ (Randomized Complete Block Design; RCBD)

ตัวอย่างที่ 2 เป็นข้อมูลจากตัวอย่างที่ 1 ที่มีการวางแผนการทดลองมาอย่างดีคือ 4 โคลน (clone) ของต้นพืชในแต่ละ genotype ที่มาจาก 6 genotypes ถูกสุ่มให้ได้รับช่วงแสงที่ต่างกัน 4 ระดับ

Genotype	v.short	short	long	v.long
A	2	3	3	4
B	3	4	5	6
C	1	2	1	2
D	1	1	2	2
E	2	2	2	2
F	1	1	2	3

เนื่องจากไม่มีซ้ำของ genotype ภายในช่วงแสง จึงไม่ใช้การวิเคราะห์แบบ Factorial ANOVA

```
photo.blk = read.table(file.choose(),header=T)
```

```
photo.blk
```

```

genotype photoperiod growth
1      A      v.short      2
2      B      v.short      3
3      C      v.short      1
4      D      v.short      1
5      E      v.short      2
6      F      v.short      1
7      A      short       3
8      B      short       4
9      C      short       2
10     D      short       1
11     E      short       2
12     F      short       1
13     A      long        3

```

```

14      B      long      5
15      C      long      1
16      D      long      2
17      E      long      2
18      F      long      2
19      A      v.long     4
20      B      v.long     6
21      C      v.long     2
22      D      v.long     2
23      E      v.long     2
24      F      v.long     3

```

```
attach(photo.blk)
```

```
names(photo.blk)
```

```
[1] "genotype"      "photoperiod" "growth"
```

```
fit = lm(growth~genotype + photoperiod)
```

```
anova(fit)
```

```
Analysis of Variance Table
```

```
Response: growth
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
genotype	5	27.875	5.5750	18.0811	7.092e-06	***
photoperiod	3	7.125	2.3750	7.7027	0.002404	**
Residuals	15	4.625	0.3083			

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

การวิเคราะห์อีกคำสั่งที่ให้ผลการวิเคราะห์ที่เหมือนกัน

```
fit2=aov(growth~genotype + photoperiod)
```

การแสดงผลของการวิเคราะห์ความแปรปรวน

```
summary(fit2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
genotype	5	27.875	5.575	18.081	7.09e-06	***
photoperiod	3	7.125	2.375	7.703	0.0024	**
Residuals	15	4.625	0.308			

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ผลการวิเคราะห์พบว่า $Pr(>F)$ มีค่า < 0.05 (0.0024) นั่นคือ อิทธิพลของช่วงแสงมีผลต่อการเจริญเติบโตของต้นพืช หรือความหมายหนึ่งคือ ช่วงแสงที่แตกต่างกันมีผลให้การเจริญเติบโตของต้นพืชต่างกัน ฉะนั้นขั้นตอนต่อไปคือการเปรียบเทียบค่าเฉลี่ยของทรีทเมนต์ว่าแต่ละทรีทเมนต์แตกต่างกันอย่างไร

การหาค่า p-value ของค่า F

```
1-pf(7.7027, 3, 15)
```

```
[1] 0.002404266
```

การเปรียบเทียบค่าเฉลี่ยของทรีทเมนต์ หลังจากพบว่าอิทธิพลของช่วงแสงมีผลต่อการเจริญเติบโตของต้นพืช

การเปรียบเทียบค่าเฉลี่ยของทรีทเมนต์ต่าง ๆ โดยใช้วิธีของ Tukey ซึ่งเป็นฟังก์ชันใน R
 TukeyHSD(fit)
 Error in UseMethod("TukeyHSD") :
 no applicable method for 'TukeyHSD' applied to an object of class "lm"

วิธีของ Tukey ใช้ได้กับการวิเคราะห์ความแปรปรวนที่ใช้คำสั่ง **aov** (โมเดลสถิติ)

TukeyHSD(fit2)

Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = growth ~ genotype + photoperiod)

\$genotype

	diff	lwr	upr	p adj
B-A	1.500000e+00	0.2243231	2.77567691	0.0170246
C-A	-1.500000e+00	-2.7756769	-0.22432309	0.0170246
D-A	-1.500000e+00	-2.7756769	-0.22432309	0.0170246
E-A	-1.000000e+00	-2.2756769	0.27567691	0.1715657
F-A	-1.250000e+00	-2.5256769	0.02567691	0.0563995
C-B	-3.000000e+00	-4.2756769	-1.72432309	0.0000187
D-B	-3.000000e+00	-4.2756769	-1.72432309	0.0000187
E-B	-2.500000e+00	-3.7756769	-1.22432309	0.0001525
F-B	-2.750000e+00	-4.0256769	-1.47432309	0.0000521
D-C	4.440892e-16	-1.2756769	1.27567691	1.0000000
E-C	5.000000e-01	-0.7756769	1.77567691	0.7943696
F-C	2.500000e-01	-1.0256769	1.52567691	0.9861980
E-D	5.000000e-01	-0.7756769	1.77567691	0.7943696
F-D	2.500000e-01	-1.0256769	1.52567691	0.9861980
F-E	-2.500000e-01	-1.5256769	1.02567691	0.9861980

\$photoperiod

	diff	lwr	upr	p adj
short-v.short	0.5000000	-0.42398727	1.423987	0.4292289
long-v.short	0.8333333	-0.09065394	1.757321	0.0841845
v.long-v.short	1.5000000	0.57601273	2.423987	0.0015096
long-short	0.3333333	-0.59065394	1.257321	0.7294779
v.long-short	1.0000000	0.07601273	1.923987	0.0318673
v.long-long	0.6666667	-0.25732061	1.590654	0.2042685

การหาค่าเฉลี่ยของทรีทเมนต์

tapply(growth,photoperiod,mean)

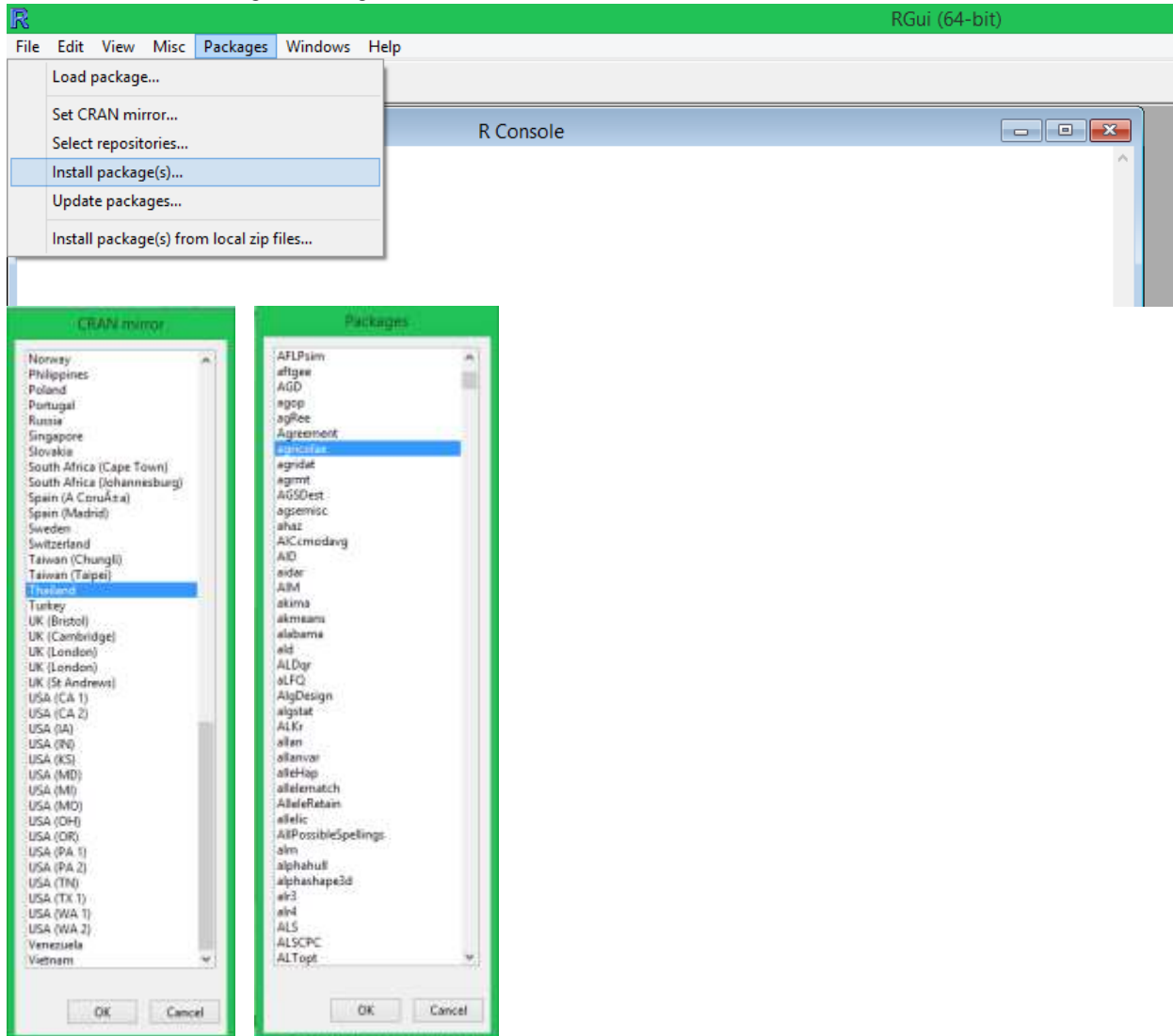
v.short	short	long	v.long
1.666667	2.166667	2.500000	3.166667

ช่วงแสง	ค่าเฉลี่ยการเจริญเติบโต (mm)
Very long	3.17 a
long	2.50 a
short	2.17 ____
Very short	1.67 ____

การใช้ package อื่นมาวิเคราะห์ร่วมกับ R

Agricolae เป็น package หนึ่งในการทดลองด้านเกษตร ที่มีวิธีการเปรียบเทียบค่าเฉลี่ยที่หลากหลายกว่าที่มีใน R ในตัวอย่างนี้ เป็นการเปรียบเทียบค่าเฉลี่ยโดยวิธี LSD (Least significant difference) ที่มีใน Agricolae มาใช้งาน

การติดตั้ง Package ที่ชื่อ agricolae ใน R



The screenshot shows the R console output for installing system binaries. The process involves downloading five ZIP files from CRAN mirrors and unpacking them. Each file's MD5 sum is checked for integrity.

```
trying URL 'http://mirrors.gnu.sc.th/pub/cran/bin/windows/contrib/3.1/x64_0-16-1.zip'
Content type 'application/zip' length 187698 bytes (183 Kb)
opened URL
downloaded 183 Kb

trying URL 'http://mirrors.gnu.sc.th/pub/cran/bin/windows/contrib/3.1/KiaR_0.4-12.zip'
Content type 'application/zip' length 319761 bytes (312 Kb)
opened URL
downloaded 312 Kb

trying URL 'http://mirrors.gnu.sc.th/pub/cran/bin/windows/contrib/3.1/apdep_0.0-63.zip'
Content type 'application/zip' length 3506718 bytes (3.4 Mb)
opened URL
downloaded 3.4 Mb

trying URL 'http://mirrors.gnu.sc.th/pub/cran/bin/windows/contrib/3.1/agricolae_1.2-1.zip'
Content type 'application/zip' length 897138 bytes (876 Kb)
opened URL
downloaded 876 Kb

package 'combinae' successfully unpacked and MD5 sums checked
package 'ap' successfully unpacked and MD5 sums checked
package 'termites' successfully unpacked and MD5 sums checked
package 'beidsi' successfully unpacked and MD5 sums checked
package 'node' successfully unpacked and MD5 sums checked
package 'kiaR' successfully unpacked and MD5 sums checked
package 'apdep' successfully unpacked and MD5 sums checked
package 'agricolae' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\Flynn\AppData\Local\Temp\RtmpOvcv\DL\downloaded_packages
```

การเปรียบเทียบค่าเฉลี่ยโดยวิธี LSD (Least significant difference)

```
compare.trt= LSD.test(growth,photoperiod,df,MSerror,group=TRUE)
```

```
compare.trt
```

```
$statistics
```

Mean	CV	MSerror	LSD
2.375	23.38011	0.3083333	0.6833208

```
$parameters
```

Df	ntr	t.value
15	4	2.13145

```
$means
```

	growth	std	r	LCL	UCL	Min	Max
long	2.500000	1.3784049	6	2.016819	2.983181	1	5
short	2.166667	1.1690452	6	1.683486	2.649847	1	4
v.long	3.166667	1.6020820	6	2.683486	3.649847	2	6
v.short	1.666667	0.8164966	6	1.183486	2.149847	1	3

```
$comparison
```

```
NULL
```

```
$groups
```

	trt	means	M
1	v.long	3.166667	a
2	long	2.500000	ab
3	short	2.166667	bc
4	v.short	1.666667	c

การเปรียบเทียบค่าเฉลี่ยโดยวิธี Duncan (: Duncan's new multiple range test)

```
compare.trt2= duncan.test(growth,photoperiod,df,MSerror,group=TRUE)
```

```
compare.trt2
```

```
$statistics
```

Mean	CV	MSerror
2.375	23.38011	0.3083333

```
$parameters
```

Df	ntr
15	4

```
$Duncan
```

	Table	CriticalRange
2	3.014325	0.6833208
3	3.159826	0.7163047
4	3.250248	0.7368026

```
$means
```

	growth	std	r	Min	Max
long	2.500000	1.3784049	6	1	5
short	2.166667	1.1690452	6	1	4
v.long	3.166667	1.6020820	6	2	6
v.short	1.666667	0.8164966	6	1	3

```
$comparison
```

```
NULL
```

```
$groups
```

	trt	means	M
1	v.long	3.166667	a
2	long	2.500000	ab
3	short	2.166667	bc
4	v.short	1.666667	c

การบันทึก Object ของผลการวิเคราะห์จาก R

การวิเคราะห์ใด ๆ ใน R หากไม่มีการบันทึกก่อนปิดโปรแกรม ข้อมูลหรือค่าที่วิเคราะห์ทั้งหมดจะหายไป การบันทึกให้บันทึก workspace โดยบันทึกเป็นชนิดไฟล์ .RData ซึ่งในกรณีนี้บันทึกเป็น rcbd.rdata และเวลาบันทึกควรพิมพ์นามสกุล .RData ด้วยทุกครั้ง เพราะการไม่พิมพ์นามสกุลเองอาจไม่ได้เป็นไฟล์ที่นำมาเปิดใน R ได้

การเรียก Object ที่บันทึกไว้มาใช้

การเรียกข้อมูลหรือผลที่วิเคราะห์แล้วบันทึกเป็น Workspace มาใช้ คำสั่งที่ใช้เรียกขึ้นมาใช้งานอีกครั้ง คือ `load("rcbd.rdata")`

การเรียก object มาใช้งาน

```
load("rcbd.rdata")
```

```
anova(fit)
```

Analysis of Variance Table

Response: growth

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
genotype	5	27.875	5.5750	18.0811	7.092e-06	***
photoperiod	3	7.125	2.3750	7.7027	0.002404	**
Residuals	15	4.625	0.3083			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

การทดลองแบบแฟคตอเรียลในแผนการทดลองแบบสุ่มตลอด (Factorial in CRD)

การทดลองแบบแฟคตอเรียลเป็นการทดลองที่มีปัจจัยที่ประกอบเป็นทรีทเมนต์มากกว่าหนึ่งปัจจัย โดยแต่ละปัจจัยจะมีระดับของปัจจัยตั้งแต่สองระดับขึ้นไป จากข้อมูลของการเจริญเติบโตของรังไข่ปลาที่ทดลองแบบ 2x2 แฟคตอเรียล ใช้แผนการทดลองแบบ CRD ที่มี 5 ซ้ำ ดังข้อมูลดิบที่ให้มาต่อไปนี้

Treatment			
a ₁ b ₁	a ₂ b ₁	a ₁ b ₂	a ₂ b ₂
1.30	1.01	0.90	0.83
2.88	1.52	1.06	0.67
2.42	1.02	0.98	0.57
2.66	1.32	1.29	0.47
2.94	1.63	1.12	0.66

โดย A = ช่วงแสง (Photoperiod): a₁ = 9 ชม. ต่อวัน และ a₂ = 14 ชม. ต่อวัน

B = อุณหภูมิ (Temperature): b₁ = 16 °C และ b₂ = 27 °C

สิ่งสำคัญเบื้องต้นคือการเตรียมข้อมูลสำหรับการวิเคราะห์ จากข้อมูลดิบที่นำมา ให้เตรียมข้อมูลในรูปแบบของ Text ไฟล์ โดยใช้ MS Excel

การอ่านข้อมูลจาก text file (txt)

```
faccrd = read.table(file.choose(),header=T)
```

```
faccrd
```

```
      photo temp ovary
1       a1    b1  1.30
2       a1    b1  2.88
3       a1    b1  2.42
4       a1    b1  2.66
5       a1    b1  2.94
6       a2    b1  1.01
7       a2    b1  1.52
8       a2    b1  1.02
9       a2    b1  1.32
10      a2    b1  1.63
11      a1    b2  0.90
12      a1    b2  1.06
13      a1    b2  0.98
14      a1    b2  1.29
15      a1    b2  1.12
16      a2    b2  0.83
17      a2    b2  0.67
18      a2    b2  0.57
19      a2    b2  0.47
20      a2    b2  0.66
```

```
attach(faccrd)
```

การวิเคราะห์ความแปรปรวน โดยการระบุตัวแปรที่จะวิเคราะห์ตามแผนการทดลอง Factorial in CRD

```
fit=aov(ovary~photo + temp + photo*temp)
```

```
summary(fit)
```

```
              Df Sum Sq Mean Sq F value    Pr(>F)
photo          1  3.081    3.081   21.704 0.000262 ***
temp           1  5.151    5.151   36.285 1.77e-05 ***
photo:temp      1  0.630    0.630    4.439 0.051268 .
Residuals     16  2.271    0.142
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tapply(ovary,photo,mean)
```

```
      a1      a2
1.755 0.970
```

```
tapply(ovary,temp,mean)
```

```
      b1      b2
1.870 0.855
```

```
tapply(ovary,list(photo,temp),mean)
```

```
      b1      b2
a1 2.44 1.07
a2 1.30 0.64
```

equivalent

```
tapply(ovary, photo:temp,mean)
```

การเปรียบเทียบค่าเฉลี่ยของ treatment combinations ต่าง ๆ ที่สะดวกโดยใช้วิธีของ Tukey

TukeyHSD(fit2)

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = ovary ~ photo + temp + photo * temp)

\$photo

	diff	lwr	upr	p adj
a2-a1	-0.785	-1.142206	-0.4277944	0.0002621

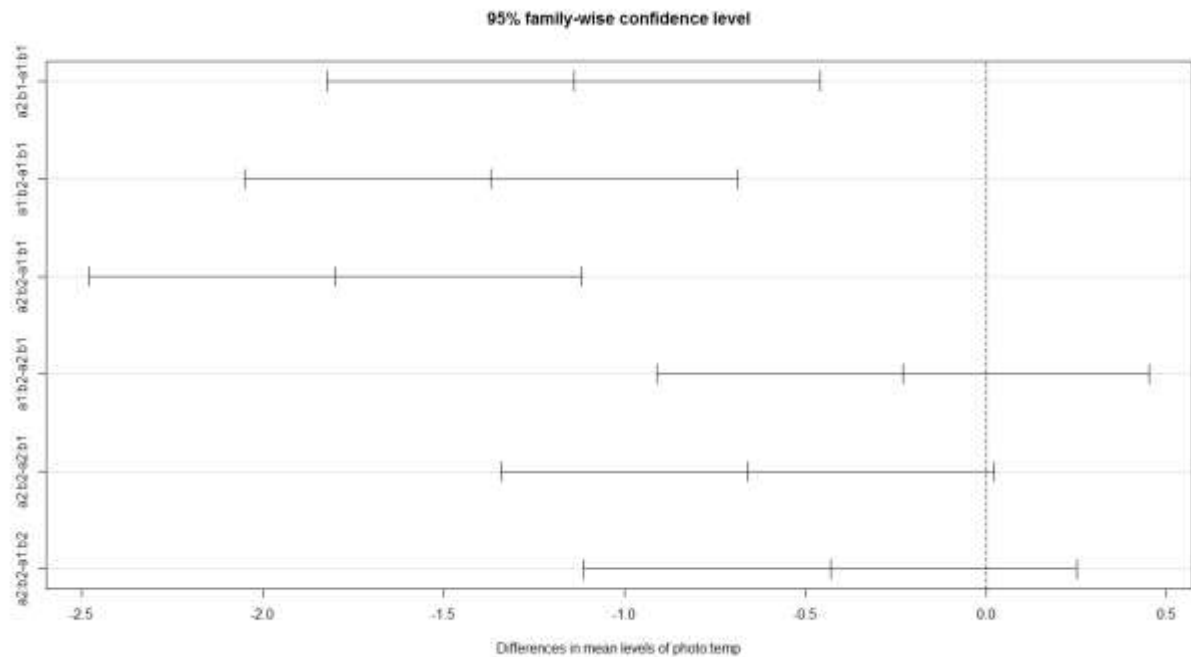
\$temp

	diff	lwr	upr	p adj
b2-b1	-1.015	-1.372206	-0.6577944	1.77e-05

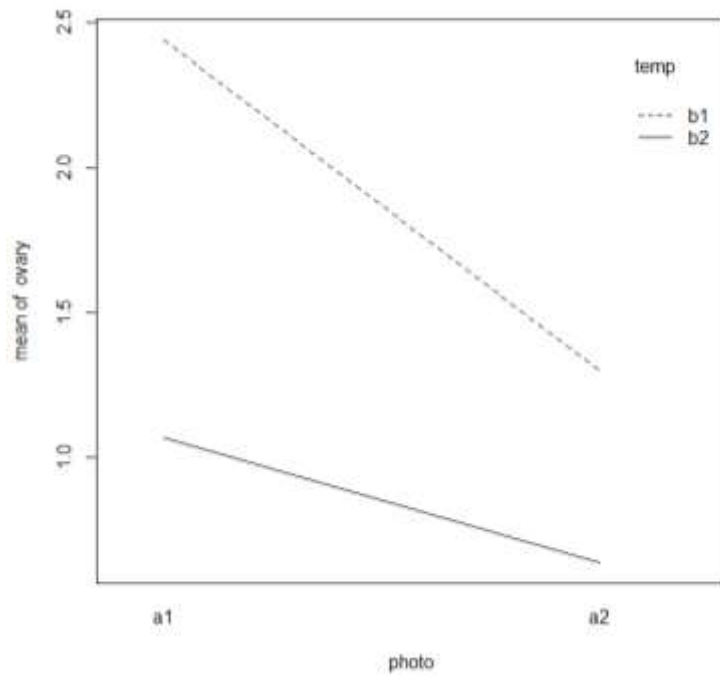
\$`photo:temp`

	diff	lwr	upr	p adj
a2:b1-a1:b1	-1.14	-1.8217697	-0.45823032	0.0010483
a1:b2-a1:b1	-1.37	-2.0517697	-0.68823032	0.0001590
a2:b2-a1:b1	-1.80	-2.4817697	-1.11823032	0.0000063
a1:b2-a2:b1	-0.23	-0.9117697	0.45176968	0.7706436
a2:b2-a2:b1	-0.66	-1.3417697	0.02176968	0.0595069
a2:b2-a1:b2	-0.43	-1.1117697	0.25176968	0.3071513

plot(TukeyHSD(fit2))



```
interaction.plot(photo,temp,ovary)
```



จากผลการวิเคราะห์ข้างบนสามารถเขียนตารางการวิเคราะห์ และตารางเปรียบเทียบค่าเฉลี่ยของทรีทเมนต์ในการทดลองได้ดังนี้

ตารางวิเคราะห์ความแปรปรวน

Source of Variation	df	SS	MS	F	P-value
Photoperiod (P)					
Temperature (T)					
P x T					
Error					
Total					

ตารางเปรียบเทียบค่าเฉลี่ยทรีทเมนต์

ทรีทเมนต์	ค่าเฉลี่ย
-----------	-----------

การเปรียบเทียบค่าเฉลี่ยโดยวิธี Duncan (: Duncan's new multiple range test)

```
compare.trt2= duncan.test(growth,photo:temp,df,MSerror,group=TRUE)
compare.trt2
```

บรรณานุกรม

- จรัญ จันทลักษณ์ และ อนันต์ชัย เชื้อนธรรม. 2540. สถิติเบื้องต้นแบบประยุกต์. ไทยวัฒนาพานิช กรุงเทพฯ. 364 น.
- สุรพล อุปติสสกุล. 2528. สถิติ การวางแผนการทดลองเบื้องต้น. 145 น.
- Cochran, W.G. and G.M. Cox. 1957. Experimental Design. 2nd edition. John Wiley and Sons, INC. 611 p.
- Crawley, M. J. 2007 The R Book. John Wiley & Sons, Ltd. ISBN: 978-0-470-51024-7. West Sussex 942 p.
- Everitt, B.S. and I. Hothorn. 2010. A Handbook of Statistical Analyses Using R. 2nd edition. Chapman & Hall/CRC. Boca Raton, FL. 348p.
- Kumar, R. 2011. Research Methodology. 3rd edition. SAGE Publications, California. 415p.

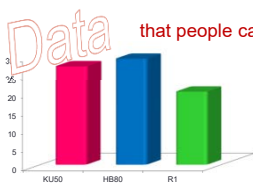

 การวิเคราะห์ทางสถิติ ด้วย **R**
R FOR STATISTICAL ANALYSES

 Asst. Prof. Dr. Piya Kittipadukul
 Department of Agronomy
 Kasetsart University (KU)
 13 February 2018



Statistics

- Collecting
- Organizing
- Analyzing
- Interpreting
- Presenting


 that people can use

Kasetsart University (KU)

Population vs. Sample

Population



Parameter
N

Sample



Statistic
n

You can never know everything but you can know sample of that
 Kasetsart University (KU)

แนวทาง และ การกระจาย

- การอธิบายถึงข้อมูลใดจะต้องอธิบายคุณสมบัติของข้อมูลสองประการ คือ แนวกลาง และ การกระจาย
 - ค่าเฉลี่ย (mean) : ค่าเฉลี่ยแบบเลขคณิต
 - ฐานนิยม (mode)
 - ค่ามัธยฐาน (median)
 - ค่าเฉลี่ยฮาร์โมนิก (harmonic mean)
 - ค่าเฉลี่ยเรขาคณิต (geometric mean)
- กรณีการแจกแจงของค่าสังเกตมีลักษณะสมมาตร (symmetric) ค่าเฉลี่ย มัธยฐาน และฐานนิยม จะมีค่าเท่ากัน
- ค่าเฉลี่ยเป็นค่าที่ไหว (sensitive) ต่อค่าสังเกตที่มาก/น้อยผิดปกติ

Kasetsart University (KU)

ค่าแสดงการกระจาย / ความแปรปรวนของข้อมูล

- ค่าพิสัย
- ค่าเบี่ยงเบนจากค่าเฉลี่ย ($\bar{x}$) คือ $(x_1 - \bar{x})$, $(x_2 - \bar{x})$, $(x_3 - \bar{x})$
- ความแตกต่างของค่าสังเกตแต่ละค่ากับค่าเฉลี่ย เรียกว่า ค่าเบี่ยงเบน (deviation) จากค่าเฉลี่ย
- ค่าวาเรียนซ์และค่าเบี่ยงเบนมาตรฐาน
- ค่าสัมประสิทธิ์ความผันแปร

Kasetsart University (KU)

Variation

- Sum of squares of deviations of x from mean
- ผลรวมของกำลังสองของค่าเบี่ยงเบนของ x จากค่าเฉลี่ย

$$\text{Sum of squares (SS)} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\text{Mean of squares (MS)} = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

(Variance, S^2)

Standard deviation, $S = \sqrt{S^2}$

ส่วนเบี่ยงเบนมาตรฐาน

$$\text{Coefficient of variation, CV} = \frac{S}{\bar{x}}, \frac{S}{\bar{x}} \times 100$$

สัมประสิทธิ์ความผันแปร

Kasetsart University (KU)

Mean of Squares

$$\text{Mean of squares (MS)} = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

(Variance, S^2)

- the sum of the actual sample values minus the sample mean, the quantities squared
- $n - 1$ is the sample size minus 1, to account for the loss of one degree of freedom from estimating by $\bar{x}$

Kasetsart University (KU)



Analysis of Variance (ANOVA)

- Treatment
- Experimental unit
- Statistical model

Kasetsart University (KU)



Treatment

- The term "treatments" used to denote the different procedures whose effects are to be measured and compared
- Factorial experiments is experiments where one factor is tested over different levels of another factor

Kasetsart University (KU)



Treatment

- Treatment / ทรีทเม้นท์ → โครงสร้างของการทดลองที่มีองอันดับแรก
- Treatment : วิธีการที่กระทำต่อสิ่งทดลองเพื่อวัดผลเปรียบเทียบตามวัตถุประสงค์
- Treatment : วิธีการ สิ่งกระทำ สิ่งกระตุ้น หรือแม้สิ่งกำหนดความแตกต่างในตัวเอง (เช่น เพศ) ที่ต้องการวัดผลจากสิ่งทดลองเพื่อเปรียบเทียบความแตกต่าง
- Treatment เป็นตัวแปร (variable) ที่ต้องการวัดตามสมมติฐาน (ตามวัตถุประสงค์)
- Treatment เป็นสิ่งที่เราเอาไปทำกับสิ่งที่ต้องการจะวัดผล (สิ่งทดลอง)

Kasetsart University (KU)



Experimental Units

- In a particular experiment, what is a unit?
- A unit is the object to which a treatment is applied, i.e. the whole unit is subject to the same treatment and different units may receive different treatments.
- The definition of units depends on the treatments, and hence on the objectives, on practicalities, on material available, etc.

Kasetsart University (KU)



Experimental Unit

- A unit is a person, animal, plant or thing which is actually studied by a researcher; the basic objects upon which the study or experiment is carried out.
- For example, a person; a monkey; a sample of soil; a pot of seedlings; a postcode area; a doctor's practice.
- Use 1 e.u. for 1 treatment = 1 replication (replicate)

Kasetsart University (KU)



Experimental Unit

- **Experimental unit** (หน่วยทดลอง) : สิ่งทดลองสิ่งเดียวหรือกลุ่มหนึ่งก็ได้ ที่จัดไว้สำหรับวัดผลของทรีตเมนต์ในการทดลอง
- **Experimental unit** : หน่วยทดลอง (หน่วยสำหรับวัด)
 - หมายถึง หน่วยที่จะทดสอบตัวอิทธิพลของ *Treatment* มีอยู่เท่าไรในการทดลอง หน่วยทดลองต้องเท่ากันหมด จะไม่เล็กบ้างใหญ่บ้าง จะหนึ่งต้องมีหน่วยวัด การทดลองในแปลงต้องมีการกำหนดขนาดของแปลง ซึ่งเรียกว่า **plot** (กำหนด *plot size*) ในทางสัตว์ เรียกว่า *pen* (คอก) หรือ *cage* (กรง)
- หน่วยทดลอง หมายถึง สิ่งหนึ่ง(กลุ่มหนึ่ง) ที่ได้รับ *Treatment* เดียวกันในครั้งใดครั้งหนึ่ง หน่วยทดลอง จะต้องเป็นหน่วยที่มีอิสระในการแสดงออกของ *Treatment* ถ้าไม่อิสระก็ไม่ใช่หน่วยทดลอง ต้องหาใหม่

Kasetsart University (KU)

CRD

Tr1	Tr2	Tr1	Tr3
Tr4	Tr3	Tr3	Tr2
Tr4	Tr1	Tr4	Tr2

Kasetsart University (KU)

RCBD

Block I	Tr1	Tr2	Tr4	Tr3
Block II	Tr4	Tr1	Tr3	Tr2
Block III	Tr3	Tr2	Tr4	Tr1

Kasetsart University (KU)

LSD

	Col1	Col2	Col3	Col4
Row1	Tr1	Tr2	Tr3	Tr4
Row2	Tr2	Tr1	Tr4	Tr3
Row3	Tr3	Tr4	Tr2	Tr1
Row4	Tr4	Tr3	Tr1	Tr2

Kasetsart University (KU)

Factorial in CRD

V1N1	V0N1	V1N0	V1N2	V0N0	V1N0
V1N0	V0N2	V0N2	V1N2	V0N0	V1N1
V0N1	V1N1	V0N1	V0N2	V1N2	V0N0

Kasetsart University (KU)

Factorial in RCBD

Blk I	V1N1	V0N2	V0N0	V1N2	V0N1	V1N0
Blk II	V1N0	V1N1	V0N2	V1N2	V0N0	V0N1
Blk III	V0N2	V1N1	V0N1	V1N0	V1N2	V0N0

Kasetsart University (KU)

Statistical Model

- แผนการทดลองแบบ **Completely Randomized Design (CRD)** มีแบบพื้นฐานทางสถิติคือ

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

- แผนการทดลองแบบ **Randomized Complete Block Design (RCBD)** มีแบบพื้นฐานทางสถิติคือ

$$Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$$

- แผนการทดลองแบบ **Latin Square (LS)** มีแบบพื้นฐานทางสถิติคือ

$$Y_{ij} = \mu + R_i + C_j + \tau_k + \varepsilon_{ijk}$$

Kasetsart University (KU)



Statistical Model

- แผนการทดลองแบบ **Factorial in RCBD** มีแบบพื้นฐานทางสถิติคือ

$$Y_{ijk} = \mu + \beta_i + A_j + B_k + AB_{jk} + \varepsilon_{ijk}$$

- แผนการทดลองแบบ **Split-plot in RCBD** มีแบบพื้นฐานทางสถิติคือ

$$Y_{ijk} = \mu + \beta_i + M_j + (\beta M)_{ij} + S_k + (MS)_{jk} + \varepsilon_{ijk}$$

Kasetsart University (KU)



Statistical Model

โดยที่

μ เป็นค่าเฉลี่ยรวม

β_i เป็นอิทธิพลของ **blocks** หรือ **replications** ของการทดลอง

τ_j เป็นอิทธิพลของ **Treatment** ของการทดลอง

R_i และ C_j เป็นอิทธิพลของ **Row** และ **Column** ตามลำดับ

A_j และ B_k เป็นอิทธิพลของปัจจัย **A** และ **B** ตามลำดับ

AB_{jk} เป็นอิทธิพลร่วมของปัจจัย **A** และ **B**

M_j และ S_k เป็นอิทธิพลของ **Main plots** และ **Subplots** ตามลำดับ

$(\beta M)_{ij}$ เป็นความแปรปรวนของ **Replication** และ **Main plots**

$(MN)_{jk}$ เป็นอิทธิพลร่วมของ **Main plots** กับ **Subplots**

ε_{ij} เป็นความคลาดเคลื่อนของการทดลอง

Kasetsart University (KU)



การวิเคราะห์แผนการทดลองแบบต่าง ๆ

- ในการวิเคราะห์แผนการทดลองจะต้องรู้แบบพื้นฐานทางสถิติ (**statistical model**) ของแผนการทดลองที่ใช้เสียก่อน

$$\text{CRD : } Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

$$\text{RCB : } Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$$

$$\text{LS : } Y_{ij} = \mu + R_i + C_j + \tau_k + \varepsilon_{ijk}$$

Kasetsart University (KU)



Statistical model

- แผนการทดลองแบบ **Factorial in RCBD** มีแบบพื้นฐานทางสถิติคือ

$$Y_{ijk} = m + \beta_i + A_j + B_k + AB_{jk} + \varepsilon_{ijk}$$

- แผนการทดลองแบบ **Split-plot in RCBD** มีแบบพื้นฐานทางสถิติคือ

$$Y_{ijk} = m + \beta_i + M_j + (\beta M)_{ij} + S_k + (MS)_{jk} + \varepsilon_{ijk}$$

Kasetsart University (KU)



แผนการทดลอง CRD

Source of variance	df	MS	F
Treatment	t - 1	Treatment MS	↖ ↗
Error	t (r - 1)	Error MS	
Total	rt - 1		

Kasetsart University (KU)



แผนการทดลอง RCBD

Source of variance	df	MS	F
Block	$r - 1$	Block MS	
Treatment	$t - 1$	Treatment MS	←
Error	$(r - 1)(t - 1)$	Error MS	
Total	$rt - 1$		

Kasetsart University (KU)

แผนการทดลอง LSD

Source of variance	df	MS	F
Row	$r - 1$	Row MS	
Column	$r - 1$	Column MS	
Treatment	$r - 1$	Treatment MS	←
Error	$(r - 1)(r - 2)$	Error MS	
Total	$r^2 - 1$		

Kasetsart University (KU)

แผนการทดลอง Factorial in RCBD

Source of variance	df	MS	F
Block	$r - 1$	Block MS	
A	$a - 1$	A MS	←
B	$b - 1$	B MS	
AB	$(a - 1)(b - 1)$	AB MS	←
Error	$(r - 1)(ab - 1)$	Error MS	
Total	$rab - 1$		

Kasetsart University (KU)

แผนการทดลอง Split-plot in RCBD

Source of variance	df	MS	F
Block	$r - 1$	Block MS	
A	$a - 1$	A MS	←
Error (a)	$(r - 1)(a - 1)$	Error a MS	
B	$b - 1$	B MS	←
AB	$(a - 1)(b - 1)$	AB MS	
Error (b)	$a(r - 1)(b - 1)$	Error b MS	
Total	$rab - 1$		

Kasetsart University (KU)

R

- A powerful *and* extensible free software environment *for statistical computing and graphics*
- A general purpose package *that includes support for a wide variety of modern statistical and graphical methods*
- It is available for most UNIX platforms, Windows and MacOS
- Part of the GNU project, and is distributed under a free software copyleft (<http://www.gnu.org/copyleft/gpl.html>)
- GNU stands for "GNU's Not Unix"

Kasetsart University (KU)

R

- The first versions of R were written by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand
- Current development is coordinated by the R Development Core Team, a committed group of volunteers

Kasetsart University (KU)

R

- *R is similar to the S language, a flexible and extensible statistical environment **originally developed in the 1980's at AT&T Bell Labs (now Lucent Technologies)***
- *Commercial software package S-PLUS™*
- *S language → the most widely-used statistical software (~ 30 years ago)*
- *S → S-PLUS → R*

Kasetsart University (KU)

**R**

- *R is powerful software **for** interacting with data*
- *can create sophisticated graphs*
- *can carryout statistical analyses*
- *can create and run simulations*

Kasetsart University (KU)

**R**

- *R is also a programming language with an extensive set of built-in functions*
- *Extend the language and write your own code to build your own statistical tools*
- *Advanced users can even incorporate functions written in other languages, such as C, C++, and Fortran*

Kasetsart University (KU)

**R**

- *An open source implementation of the S language*
- **Free**
- *High-quality statistical software*
- *A first-rate tool for professional statistician*

Kasetsart University (KU)

**Go Get R**

- *Installing R on your computer is simple*
- *R website: <http://cran.r-project.org>
or www.r-project.org*

Kasetsart University (KU)

**Case sensitive**

- *Unlike SAS, R is case sensitive*
- *x <- 1:3*
- *X <- seq(2, 4)*
- *x*
- *[1] 1 2 3*
- *X*
- *[1] 2 3 4*

Kasetsart University (KU)



Object

- Almost everything in R *is an object*
- An object is simply something that R can operate on
- Common objects include vectors, matrices, arrays, factors, dataframes (*akin to datasets in SAS*), lists, and functions

Kasetsart University (KU)



Running R

- `>` character is the prompt, and commands are executed once the user presses the RETURN key
- `<-` is the assignment operator
- `=` is another assignment operator
is generally recommended for the specification of options for functions rather than assignment
- `help("c-")` or `? "c-"` for more information
- The `$` operator allows direct access to objects within a dataframe
- The `attach()` command can be used to make objects within a dataset available in the global workspace from that point forward

Kasetsart University (KU)



Entering data

- Entering data using `c`
- Entering data using `scan`
- Entering data using `read.table`
- Reading in data using `file.choose`
- Reading in data from an Excel file

Kasetsart University (KU)



Entering data using c

- The easiest way to enter small data sets is with the function `c`
- ```
> glucose = c(81, 85, 93, 93, 99, 76, 75, 84, 78, 84, 81, 82, 89,
+ 81, 96, 82, 74, 70, 84, 86, 80, 70, 131, 75, 88, 102, 115,
+ 89, 82, 79, 106)
```
- Concatenates numbers (or vectors) together
  - The `' '` symbols, which are prompts that indicates R is waiting for a command to be completed
  - press [Enter] to continue a command on the next line

Kasetsart University (KU)



## Entering data using scan

- The `scan` function reads in data separated by "white space", spaces, tabs, and newline characters
  - If you press [Enter] on a blank line, you end the input
- ```
> glucose = scan()
1: 81 85 93 93 99 76 75 84 78 84 81 82 89 81
15: 96 82 74 70 84 86 80 70 131 75 88 102
27: 115 89 82 79 106
32:
Read 31 items
```

Kasetsart University (KU)



Entering data using read.table

- The easiest to read into R are the ASCII formatted files, or plain text files
- If you have copied this file to your hard drive and if you have changed directories through the File menu so that your working directory contains this data file
- `> x = read.table("filename.txt", header = T)`
- The R commands above create `x`, an object known in R as a data frame

Kasetsart University (KU)



Entering data using read.table

- The default behavior of R is to read in data frames without headers
 - A data frame is a data matrix, but can include both categorical and numerical variables
 - Use the command `str` to check the structure of the data frame
- ```
> str(x)
'data.frame': 31 obs. of 1 variable:
 $ glucose: int 81 85 93 93 99 76 75 84 78 84 ...
```

Kasetsart University (KU)



### Entering data using read.table

- After reading in a data frame, R will recognize the name `x` as the name of the entire data frame, but R will not recognize the variable `glucose`
  - There are two ways to get around this. First, the variables of each data frame may be accessed using the `$` operator
- ```
> x$glucose
[1] 81 85 93 93 99 76 75 84 78 84 81 82 89 81 96 82 74 70 84
[20] 86 80 70 131 75 88 102 115 89 82 79 106
```

Kasetsart University (KU)



Entering data using read.table

- Alternatively, if you use the `attach` command, you can refer to each variable in `x` by name without the dollar sign
- ```
> attach(x)
> glucose
[1] 81 85 93 93 99 76 75 84 78 84 81 82 89 81 96 82 74 70 84
[20] 86 80 70 131 75 88 102 115 89 82 79 106
```
- You only need to type in the `attach` command once per session--once you have done it, R will know that `glucose` means the same thing as `x$glucose` until you quit

Kasetsart University (KU)



### Reading in data using file.choose

- `file.choose` will open up a dialog box that you can use to move through your folders to find the data file you want to read in
- ```
x = read.table(file.choose(),header=T)

x <- read.table('clipboard',header=T)
```

Kasetsart University (KU)



Reading in data from an Excel file

- R does not read in data directly from an Excel file, but you can use Excel to export data in a format that R can read
 - Save the spread sheet as a CSV file, for comma-separated-variable file. This format is a plain text file, but there will be commas instead of spaces between variable values
 - using `read.table` with the additional argument `sep=","`
- ```
> name = read.table(filename.csv", header = T, sep=",")
```

Kasetsart University (KU)



### Keep in mind

- Packages within R are created by individuals or small groups, and not endorsed by the R core group
- As a result, they do not necessarily undergo the same level of testing and quality assurance that the core R system does

Kasetsart University (KU)



**Case 1**

| v.short | short | long | v.long |
|---------|-------|------|--------|
| 2       | 3     | 3    | 4      |
| 3       | 4     | 5    | 6      |
| 1       | 2     | 1    | 2      |
| 1       | 1     | 2    | 2      |
| 2       | 2     | 2    | 2      |
| 1       | 1     | 2    | 3      |

Kasetsart University (KU)

**Command**

- `photo = read.table(file.choose(),header=T)`
- `photo`
- `attach(photo)`
- `names(photo)`
- `tapply(growth,photoperiod,mean)`

Kasetsart University (KU)

**Model**

- `growth ~ photoperiod`
- `fit = lm(growth~photoperiod)`
- `anova(fit)`

Kasetsart University (KU)

**Analysis of Variance Table**

Response: growth

|             | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-------------|----|--------|---------|---------|--------|
| photoperiod | 3  | 7.125  | 2.375   | 1.4615  | 0.2551 |
| Residuals   | 20 | 32.500 | 1.625   |         |        |

- critical values of Fisher's F
- `qf(0.95,3,20)`

Kasetsart University (KU)

**Case 2**

| Genotype | v.short | short | long | v.long |
|----------|---------|-------|------|--------|
| A        | 2       | 3     | 3    | 4      |
| B        | 3       | 4     | 5    | 6      |
| C        | 1       | 2     | 1    | 2      |
| D        | 1       | 1     | 2    | 2      |
| E        | 2       | 2     | 2    | 2      |
| F        | 1       | 1     | 2    | 3      |

Kasetsart University (KU)

**RCBD**

- `photo.blk = read.table(file.choose(),header=T)`
- `photo.blk`
- `attach(photo.blk)`
- `names(photo.blk)`
- `fit = lm(growth~genotype + photoperiod)`
- `anova(fit)`
- `TukeyHSD(fit2)`

Kasetsart University (KU)



### Result

Response: growth

|             | Df | Sum Sq | Mean Sq | F value | Pr(>F)        |
|-------------|----|--------|---------|---------|---------------|
| genotype    | 5  | 27.875 | 5.5750  | 18.0811 | 7.092e-06 *** |
| photoperiod | 3  | 7.125  | 2.3750  | 7.7027  | 0.002404 **   |
| Residuals   | 15 | 4.625  | 0.3083  |         |               |

\*\*\*

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Kasetsart University (KU)



### Mean Comparison

| ช่วงแสง    | ค่าเฉลี่ยการเจริญเติบโต (mm) |
|------------|------------------------------|
| Very long  | 3.17 a                       |
| long       | 2.50 a                       |
| short      | 2.17 ____                    |
| Very short | 1.67 ____                    |

Kasetsart University (KU)



### Asst. Prof. Dr. Piya Kittipadakul

- Head of Department of Agronomy
- Faculty of Agriculture, Kasetsart University
- kittipadakul@yahoo.com
- Facebook: [Thai Cassava](#)
- Facebook Page: [Thai Cassava - มั่นสำปะหลังไทย](#)

Thank you

Kasetsart University (KU)

