

# ZeaMays Data

coop711

2018-03-31

## Data

### HistData

ZeaMays Data 는 HistData package 에 있음.

```
library(knitr)
library(pander)
# install.packages("HistData", repos = "http://cran.rstudio.com")
library(HistData)
ZeaMays
```

```
##   pair pot  cross  self  diff
## 1    1  1 23.500 17.375  6.125
## 2    2  1 12.000 20.375 -8.375
## 3    3  1 21.000 20.000  1.000
## 4    4  2 22.000 20.000  2.000
## 5    5  2 19.125 18.375  0.750
## 6    6  2 21.500 18.625  2.875
## 7    7  3 22.125 18.625  3.500
## 8    8  3 20.375 15.250  5.125
## 9    9  3 18.250 16.500  1.750
## 10   10  3 21.625 18.000  3.625
## 11   11  3 23.250 16.250  7.000
## 12   12  4 21.000 18.000  3.000
## 13   13  4 22.125 12.750  9.375
## 14   14  4 23.000 15.500  7.500
## 15   15  4 12.000 18.000 -6.000
```

```
str(ZeaMays)
```

```
## 'data.frame':   15 obs. of  5 variables:
## $ pair : int  1 2 3 4 5 6 7 8 9 10 ...
## $ pot : Factor w/ 4 levels "1","2","3","4": 1 1 1 2 2 2 3 3 3 3 ...
## $ cross: num  23.5 12 21 22 19.1 ...
## $ self : num  17.4 20.4 20 20 18.4 ...
## $ diff : num  6.12 -8.38 1 2 0.75 ...
```

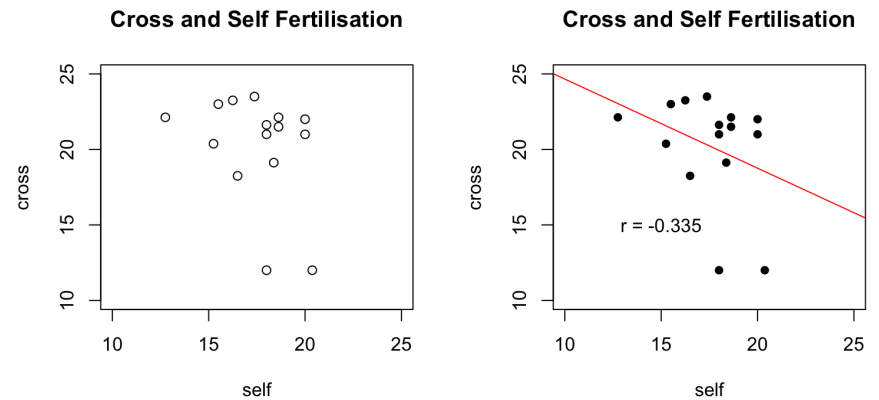
## Competition

산점도로 Competition 현상 확인.

```
par(mfrow = c(1, 2))
plot(cross ~ self,
     data = ZeaMays,
     xlim = c(10, 25),
     ylim = c(10, 25))
title(main = "Cross and Self Fertilisation")
plot(cross ~ self,
     data = ZeaMays,
     pch = 16,
     xlim = c(10, 25),
     ylim = c(10, 25))
title(main = "Cross and Self Fertilisation")
abline(lm(cross ~ self, data = ZeaMays)$coef,
       col = "red")
cor(ZeaMays$self, ZeaMays$cross)
```

```
## [1] -0.3347553
```

```
text(x = 15, y = 15,
     labels = paste("r =", round(cor(ZeaMays$cross, ZeaMays$self), digits = 3)))
```

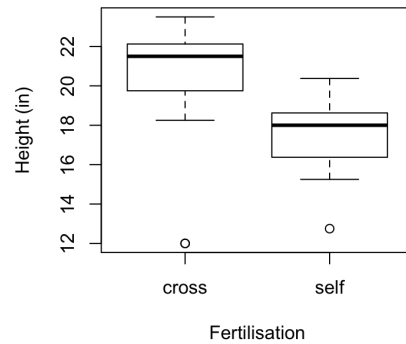
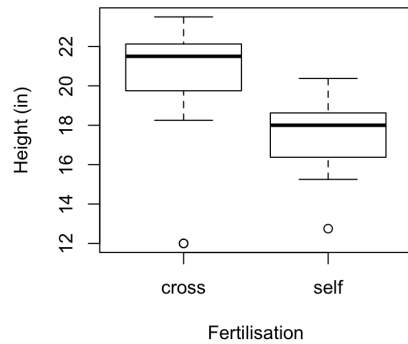


```
par(mfrow = c(1, 1))
```

## Boxplot

`boxplot()` 으로 비교하면 이 현상을 파악하기 어려움. 같은 결과를 갖는 두 가지 `boxplot()` 작성 코드 비교.

```
par(mfrow = c(1,2))
boxplot(ZeaMays[c("cross", "self")],
        ylab = "Height (in)",
        xlab = "Fertilisation")
boxplot(ZeaMays$cross, ZeaMays$self,
        names = c("cross", "self"),
        ylab = "Height (in)",
        xlab = "Fertilisation")
```



## Paired 1-sample t-test

쌍으로 키우고 있으므로 성장의 차이는 paired one-sample t-test

```
t.test(x = ZeaMays$cross, y = ZeaMays$self,
       paired = T)
```

```
##
## Paired t-test
##
## data: ZeaMays$cross and ZeaMays$self
## t = 2.148, df = 14, p-value = 0.0497
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.003899165 5.229434169
## sample estimates:
## mean of the differences
##                2.616667
```

또는 둘 사이의 차이인 `diff` 대하여 one-sample t-test 를 수행하여도 동일한 결과. 이 때는 모든 매개변수는 디폴트 값으로 충족됨에 유의.

```
t.test(ZeaMays$diff)
```

```
##
## One Sample t-test
##
## data: ZeaMays$diff
## t = 2.148, df = 14, p-value = 0.0497
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.003899165 5.229434169
## sample estimates:
## mean of x
##  2.616667
```

## Tests of Normality

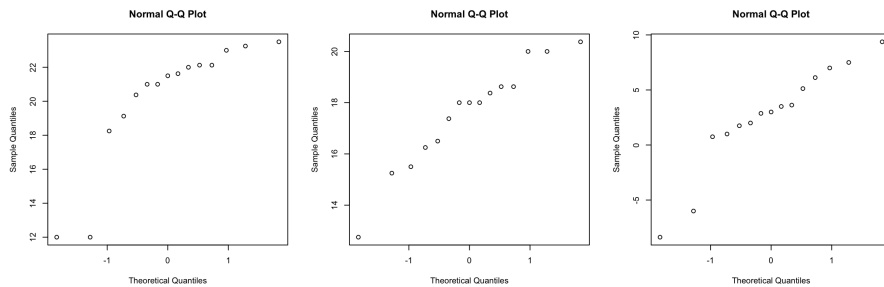
정규성에 대한 가정은 `ad.test()` 로 수행

```
library(nortest)
sapply(ZeaMays[c("cross", "self", "diff")],
       FUN = ad.test)

##           cross
## statistic 1.475455
## p.value   0.0005056947
## method    "Anderson-Darling normality test"
## data.name "X[[i]]"
##           self
## statistic 0.3546936
## p.value   0.4122312
## method    "Anderson-Darling normality test"
## data.name "X[[i]]"
##           diff
## statistic 0.6175125
## p.value   0.08757741
## method    "Anderson-Darling normality test"
## data.name "X[[i]]"
```

cross 자료에 대한 정규성은 매우 의심되는 수준. `qqnorm()` 으로 파악.

```
par(mfrow = c(1,3))
sapply(ZeaMays[c("cross", "self", "diff")],
       FUN = qqnorm)
```



```
## cross      self      diff
## x Numeric,15 Numeric,15 Numeric,15
## y Numeric,15 Numeric,15 Numeric,15
```

```
par(mfrow = c(1, 1))
```

## Rank Tests

`t.test()` 의 대안으로 `wilcox.test()` 수행. 먼저 signed ranks를 구하면.

```
sign(ZeaMays$diff) * rank(abs(ZeaMays$diff))
```

```
## [1] 11 -14 2 4 1 5 7 9 3 8 12 6 15 13 -10
```

1에서 15까지 자연수의 합은 120 이므로 통계량은 양의 부호 순위합임.

```
wilcox.test(ZeaMays$cross, ZeaMays$self)
```

```
## Warning in wilcox.test.default(ZeaMays$cross, ZeaMays$self): cannot compute
## exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: ZeaMays$cross and ZeaMays$self
## W = 185.5, p-value = 0.002608
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(ZeaMays$diff)
```

```
##
## Wilcoxon signed rank test
##
## data: ZeaMays$diff
## V = 96, p-value = 0.04126
## alternative hypothesis: true location is not equal to 0
```

따라서, cross 와 self 가 성장률이 동일하다는 가설은 기각됨. 화분 간의 차이가 있는지 분산분석으로 살펴보면

## anova()

```
kable(anova(lm(diff ~ pot, data = ZeaMays)))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pot	3	44.69245	14.89748	0.6138755	0.6200915
Residuals	11	266.94714	24.26792	NA	NA

따라서, 화분 간의 차이는 없다고 결론.