ZeaMays Data

coop711

2015년 3월 30일

- ZeaMays Data 는 "HistData" package 에 있음.

```
install.packages("HistData", repos="http://cran.rstudio.com")
```

```
##
## The downloaded binary packages are in
##   /var/folders/_h/tg1th9bd4h98rjjb5vy9gn3m0000gn/T//RtmpVR9Jf9/downloaded_pac
kages
```

```
library(HistData)
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 ...
```

```
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
```

```
attach(ZeaMays)
```
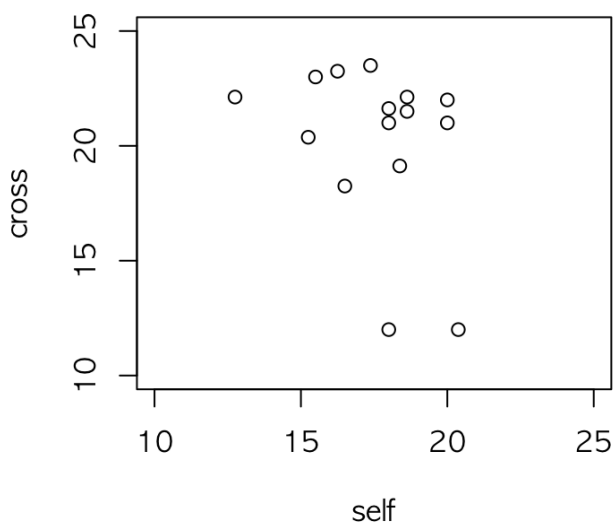
- Competition 현상 확인.

```
par(mfrow=c(1,2))
plot(x=self, y=cross, xlim=c(10, 25), ylim=c(10, 25))
title(main="Cross and Self Fertilisation")
plot(x=self, y=cross, pch=16, xlim=c(10, 25), ylim=c(10, 25))
title(main="Cross and Self Fertilisation")
abline(lsfit(x=self, y=cross)$coef, col="red")
cor(self, cross)
```
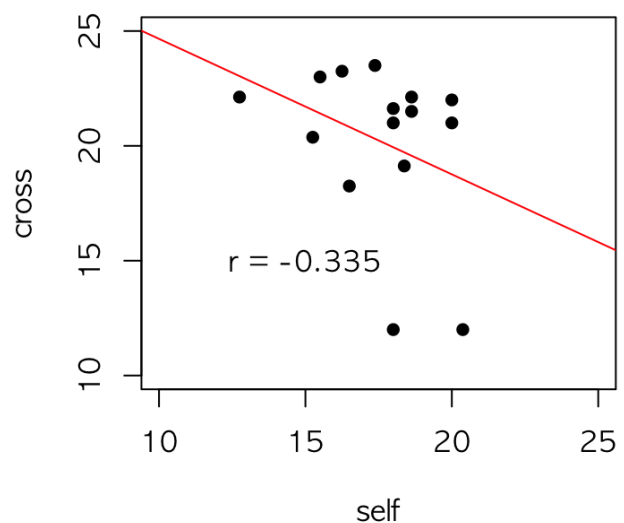
```
## [1] -0.3347553
```

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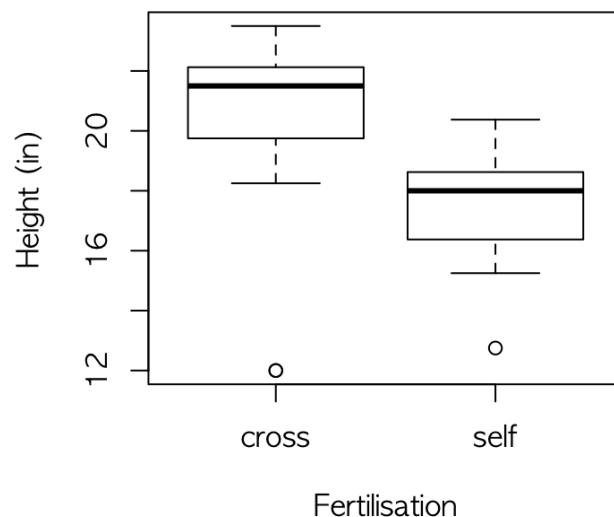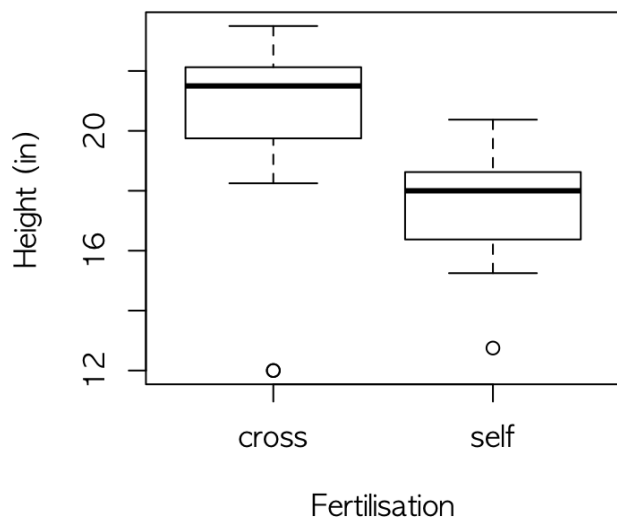


```
par(mfrow=c(1,1))
```

- 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="Fertilisatio
n")
boxplot(cross, self, names=c("cross", "self"), ylab="Height (in)", xlab="Fertil
isation")
```

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

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

```
##
##  Paired t-test
##
## data:  cross and 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(diff)
```

```
##
##  One Sample t-test
##
## data:  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
```

- 정규성에 대한 가정은 ad.test로 수행

```
library(nortest)
lapply(list(cross=cross, self=self, diff=diff), ad.test)
```

```
## $cross
##
##   Anderson-Darling normality test
##
## data:  X[[1L]]
## A = 1.4755, p-value = 0.0005057
##
##
## $self
##
##   Anderson-Darling normality test
##
## data:  X[[2L]]
## A = 0.3547, p-value = 0.4122
##
##
## $diff
##
##   Anderson-Darling normality test
##
## data:  X[[3L]]
## A = 0.6175, p-value = 0.08758
```

- cross 자료에 대한 정규성은 매우 의심되는 수준. qqnorm으로 파악.

```
par(mfrow=c(1,3))
lapply(list(cross=cross, self=self, diff=diff), qqnorm)
```
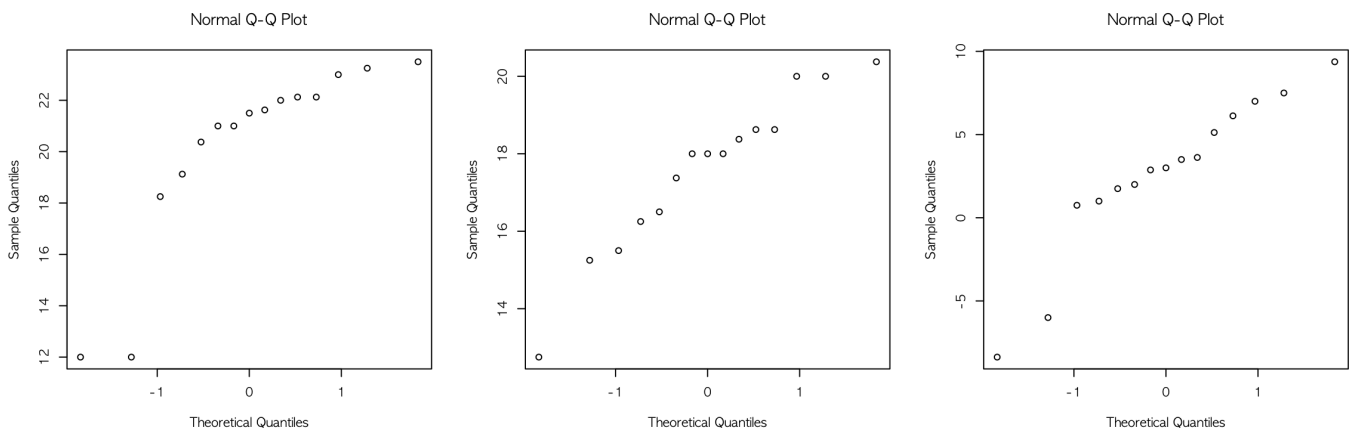
```
## $cross
## $cross$x
##  [1]  1.8339146 -1.8339146 -0.3406948  0.3406948 -0.7279133  0.0000000
##  [7]  0.5244005 -0.5244005 -0.9674216  0.1678940  1.2815516 -0.1678940
## [13]  0.7279133  0.9674216 -1.2815516
##
## $cross$y
##  [1] 23.500 12.000 21.000 22.000 19.125 21.500 22.125 20.375 18.250 21.625
## [11] 23.250 21.000 22.125 23.000 12.000
##
##
## $self
## $self$x
##  [1] -0.3406948  1.8339146  0.9674216  1.2815516  0.3406948  0.5244005
##  [7]  0.7279133 -1.2815516 -0.5244005 -0.1678940 -0.7279133  0.0000000
## [13] -1.8339146 -0.9674216  0.1678940
##
## $self$y
##  [1] 17.375 20.375 20.000 20.000 18.375 18.625 18.625 15.250 16.500 18.000
## [11] 16.250 18.000 12.750 15.500 18.000
##
##
## $diff
## $diff$x
##  [1]  0.7279133 -1.8339146 -0.7279133 -0.3406948 -0.9674216 -0.1678940
##  [7]  0.1678940  0.5244005 -0.5244005  0.3406948  0.9674216  0.0000000
## [13]  1.8339146  1.2815516 -1.2815516
##
## $diff$y
##  [1]  6.125 -8.375  1.000  2.000  0.750  2.875  3.500  5.125  1.750  3.625
## [11]  7.000  3.000  9.375  7.500 -6.000
```

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

- t.test의 대안으로 wilkox.test 수행

```
wilcox.test(cross, self, paired=T)
```

```
##
##   Wilcoxon signed rank test
##
## data:  cross and self
## V = 96, p-value = 0.04126
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(diff)
```

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

- cross와 self 가 성장률이 동일하다는 가설은 기각됨.