

# fruit flies

Chris Parrish

January 19, 2016

fruit flies

reference:

- Cannon, et al., Stat2, chapter 05, examples 5.1-5.6, 5.10-5.11
- Cannon, R Manual, chapter 5

Import the data.

```
data <- read.csv("FruitFlies.csv", header=TRUE)
head(data, 4)
```

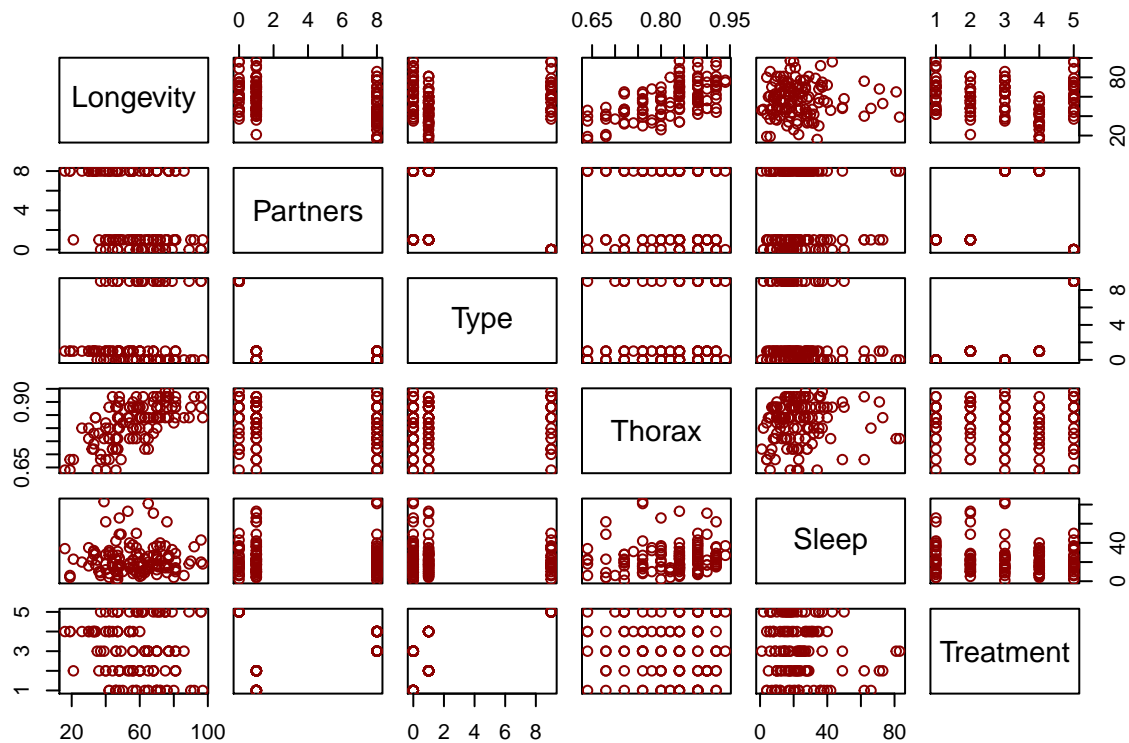
```
##   ID Partners Type Longevity Thorax Sleep Treatment
## 1  1      8    0      35    0.64    22 8 pregnant
## 2  2      8    0      37    0.68     9 8 pregnant
## 3  3      8    0      49    0.68    49 8 pregnant
## 4  4      8    0      46    0.72     1 8 pregnant
```

```
dim(data)
```

```
## [1] 125  7
```

Scatterplot matrix.

```
pairs(~ Longevity + Partners + Type + Thorax + Sleep + Treatment, data=data,
      col="darkred")
```



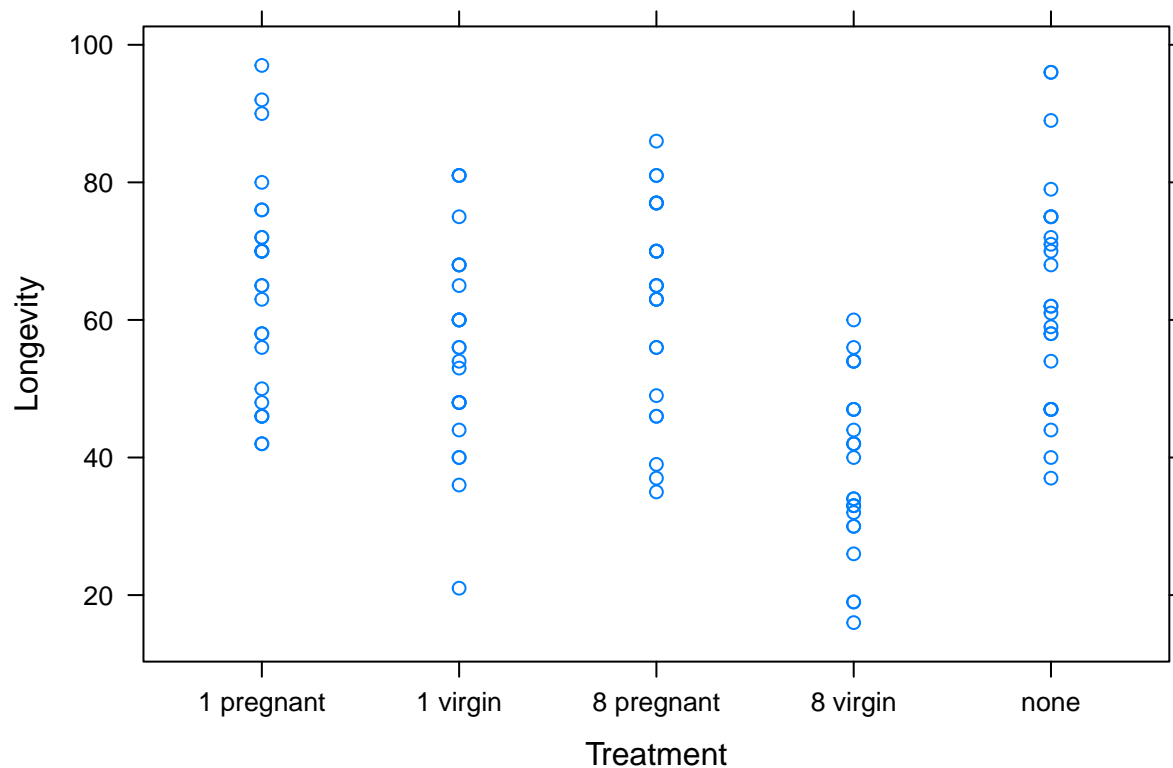
Variable *Treatment* is a Factor.

```
str(data)
```

```
## 'data.frame': 125 obs. of 7 variables:
## $ ID : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Partners : int 8 8 8 8 8 8 8 8 8 8 ...
## $ Type : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Longevity: int 35 37 49 46 63 39 46 56 63 65 ...
## $ Thorax : num 0.64 0.68 0.68 0.72 0.72 0.76 0.76 0.76 0.76 ...
## $ Sleep : int 22 9 49 1 23 83 23 15 9 81 ...
## $ Treatment: Factor w/ 5 levels "1 pregnant", "1 virgin",...: 3 3 3 3 3 3 3 3 3 3 ...
```

Use Lattice graphics to view the data.

```
library(lattice)
xyplot(Longevity ~ Treatment, data=data)
```



Group statistics.

```
n <- with(data, tapply(Longevity, Treatment, length))
mean <- with(data, round(tapply(Longevity, Treatment, mean), 3))
sd <- with(data, round(tapply(Longevity, Treatment, sd), 3))
idx <- c(5, 1, 3, 2, 4) # idx orders the rows in the table
fruitfly.statistics <- cbind(n, mean, sd)[idx, ]
fruitfly.statistics
```

```
##           n mean    sd
```

```
## none      25 63.56 16.452
## 1 pregnant 25 64.80 15.652
## 8 pregnant 25 63.36 14.540
## 1 virgin   25 56.76 14.928
## 8 virgin   25 38.72 12.102
```

ANOVA with aov.

```
fruitfly.aov <- aov(Longevity ~ Treatment, data=data)
fruitfly.aov
```

```
## Call:
## aov(formula = Longevity ~ Treatment, data = data)
##
## Terms:
##           Treatment Residuals
## Sum of Squares  11939.28 26313.52
## Deg. of Freedom      4      120
##
## Residual standard error: 14.80808
## Estimated effects may be unbalanced
```

```
options(show.signif.stars=FALSE)
summary(fruitfly.aov)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Treatment    4  11939   2984.8   13.61 3.52e-09
## Residuals   120  26314    219.3
```

SS and degrees of freedom for ANOVA.

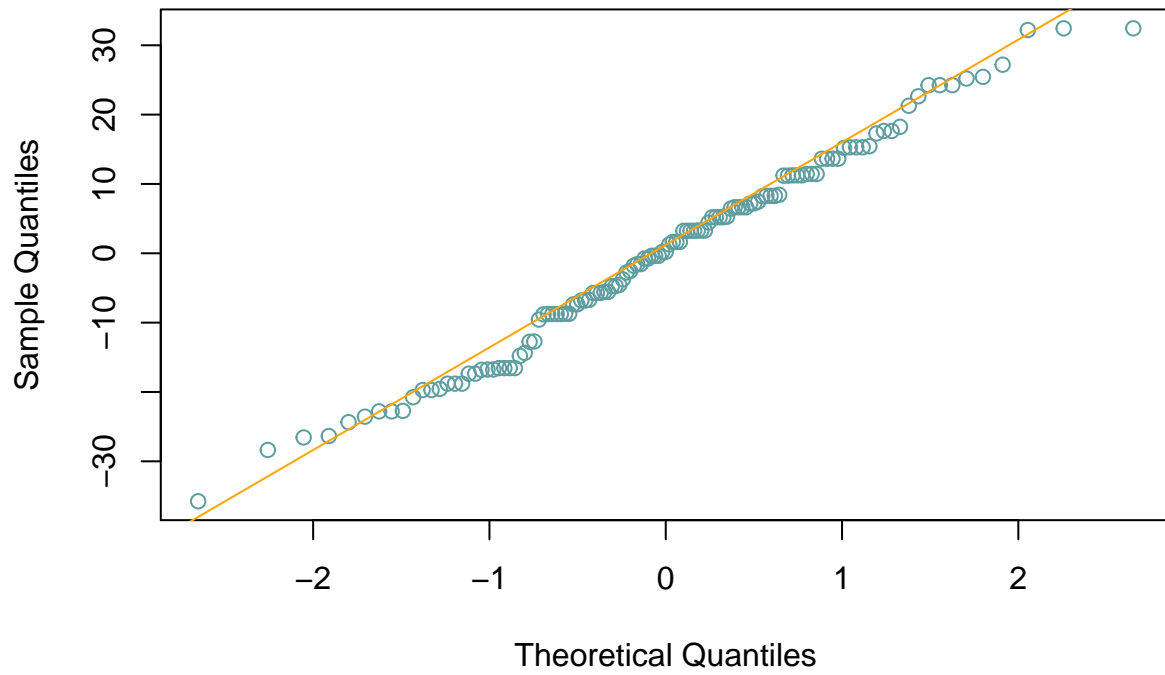
*Observed - Grand Mean = Group Effect + Residual*

$df_{Total} = n - 1 = df_{Groups} + df_{Error} = (k - 1) + (n - k)$

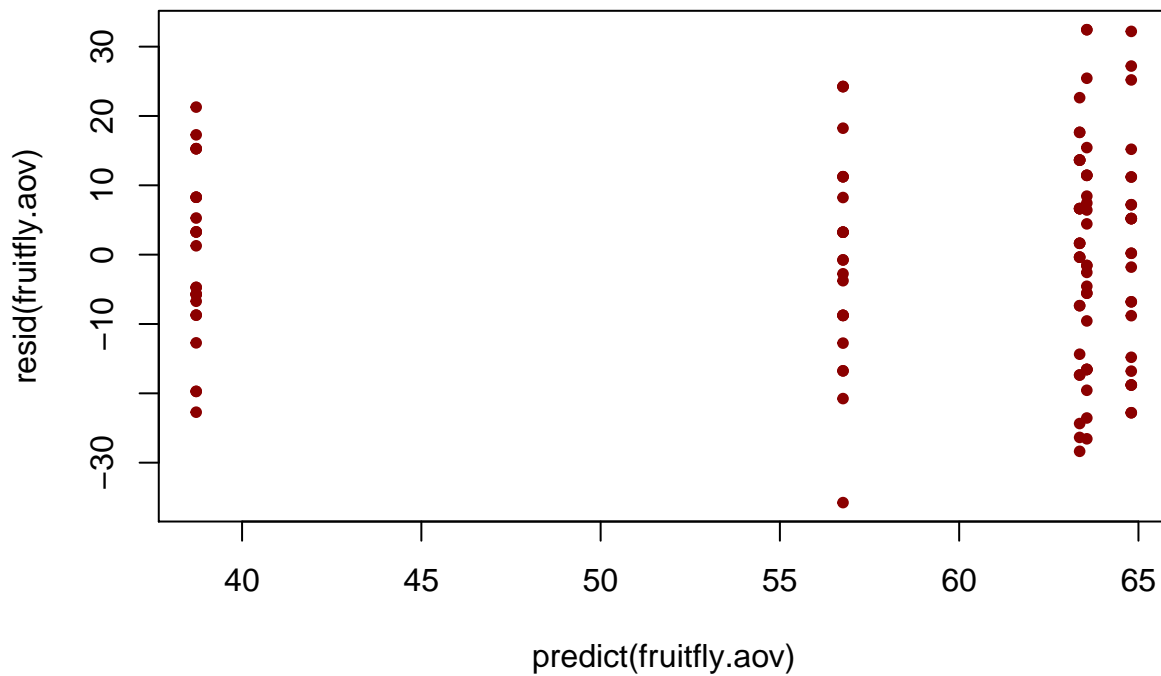
Residuals.

```
qqnorm(resid(fruitfly.aov), col="cadetblue")
qqline(resid(fruitfly.aov), col="orange")
```

## Normal Q-Q Plot



```
plot(predict(fruitfly.aov), resid(fruitfly.aov),  
     pch=20, col="darkred")
```



```
std.dev <- fruitfly.statistics[, 3]  
std.dev
```

```
##      none 1 pregnant 8 pregnant   1 virgin   8 virgin
```

```
##      16.452      15.652      14.540      14.928      12.102
```

```
ratio <- max(std.dev) / min(std.dev)
ratio
```

```
## [1] 1.359445
```

Fisher's Least Significant Difference (= LSD).

$$LSD = t^* \sqrt{MSE \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

If a difference of means is larger than the *LSD* of those means then the associated CI does not contain 0, so the means are significantly different.

Test for the difference of two means.

$$CI : \bar{y}_{8v} - \bar{y}_{none} = \pm t^* \sqrt{MSE \left( \frac{1}{n_{8v}} + \frac{1}{n_{none}} \right)}$$

```
y.bar.8v <- fruitfly.statistics[5, 2]
y.bar.none <- fruitfly.statistics[1, 2]
point.estimate <- y.bar.8v - y.bar.none
alpha <- 0.05
df <- 120
t.star <- qt(c(alpha/2, 1 - alpha/2), df=df)
mse <- 219.3
n.8v <- n.none <- 25
se <- sqrt(mse * (1 / n.8v + 1 / n.none))
ci <- point.estimate + t.star * se
ci
```

```
## [1] -33.13305 -16.54695
```

Note that the right endpoint of the CI reported in the text (p.254) is not correct.

Test all possible pairs of means.

Each sample size is 25, so compare the absolute value of each difference of means with Fisher's LSD. If the difference of means is larger than Fisher's LSD, then the means are significantly different (because the associated CI will not contain 0).

Which pairs of means are significantly different?

```
LSD <- t.star[2] * se
LSD
```

```
## [1] 8.293049
```

```
diffs <- outer(mean, t(mean), "--")
diffs
```

```

## , , 1 pregnant
##
##          [,1]
## 1 pregnant  0.00
## 1 virgin   -8.04
## 8 pregnant -1.44
## 8 virgin   -26.08
## none      -1.24
##
## , , 1 virgin
##
##          [,1]
## 1 pregnant  8.04
## 1 virgin    0.00
## 8 pregnant  6.60
## 8 virgin   -18.04
## none        6.80
##
## , , 8 pregnant
##
##          [,1]
## 1 pregnant  1.44
## 1 virgin   -6.60
## 8 pregnant  0.00
## 8 virgin   -24.64
## none        0.20
##
## , , 8 virgin
##
##          [,1]
## 1 pregnant 26.08
## 1 virgin   18.04
## 8 pregnant 24.64
## 8 virgin    0.00
## none       24.84
##
## , , none
##
##          [,1]
## 1 pregnant  1.24
## 1 virgin   -6.80
## 8 pregnant -0.20
## 8 virgin   -24.84
## none        0.00

```

Effect plot: without interaction.

```

fruitflies.lm1 <- lm(Longevity ~ Treatment, data=data)
library(alr4)
plot(allEffects(fruitflies.lm1))

```

Treatment effect plot

