q2 example

Daniel Detore

```
CropYield = data.frame(read.csv("./Crop yield.csv"))
CropYield$block = as.factor(CropYield$block)
CropYield$density = as.factor(CropYield$density)
CropYield$fertilizer = as.factor(CropYield$fertilizer)
```

1

1.1

H_0 : ‘Block’ and ‘Density’ have no association.

H_a : ‘Block’ and ‘Density’ have some association.

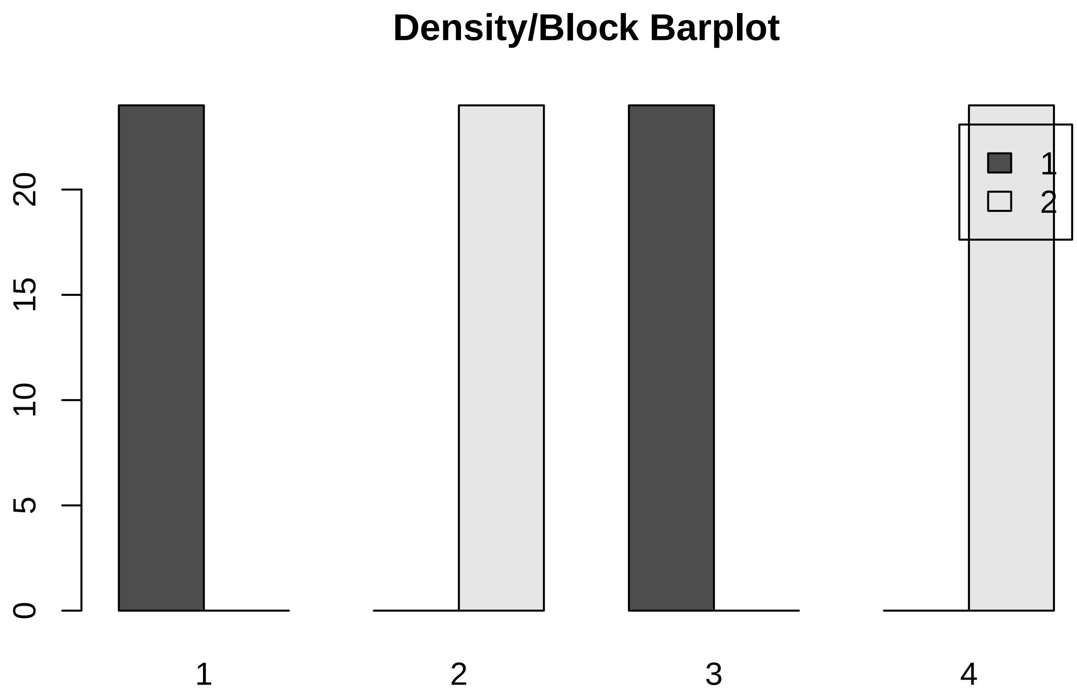
1.2

```
table = table(CropYield$density, CropYield$block); table
```

```
##
##      1  2  3  4
##   1 24  0 24  0
##   2  0 24  0 24
```

##1.3

```
barplot(table, legend=TRUE, beside=TRUE, main='Density/Block Barplot')
```



They are of different patterns. Blocks 1 and 3 only have density 1, while blocks 2 and 4 only have density 2.

1.4

The value of each inner cell $e_{i,j} = \frac{o_{i,+}o_{+,j}}{o_{+,+}}$.

```
o_ip <- rowSums(table)
o_pj <- colSums(table)
o_pp <- sum(table)
e <- outer(o_ip, o_pj)/o_pp
rownames(e) = c("Density 1", "Density 2")
colnames(e) = c("Block 1", "Block 2", "Block 3", "Block 4")
e
```

```
##           Block 1 Block 2 Block 3 Block 4
## Density 1      12      12      12      12
## Density 2      12      12      12      12
```

1.5

```
chisq.test(table)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table
## X-squared = 96, df = 3, p-value < 2.2e-16
```

The testing statistic

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}}$$

is observed as

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{i,j}-e_{i,j})^2}{e_{i,j}} = 96$$

H_0 is χ^2_k where $k = (r-1)(c-1) = (2-1)(4-1) = 3$. The p-value is less than 2.2×10^{-16}.

1.6

In this case we must reject H_0 because $2.2 \times 10^{-16} < \alpha = 0.1$. Thus we assume H_a, that 'Block' and 'Density' have some association.

2

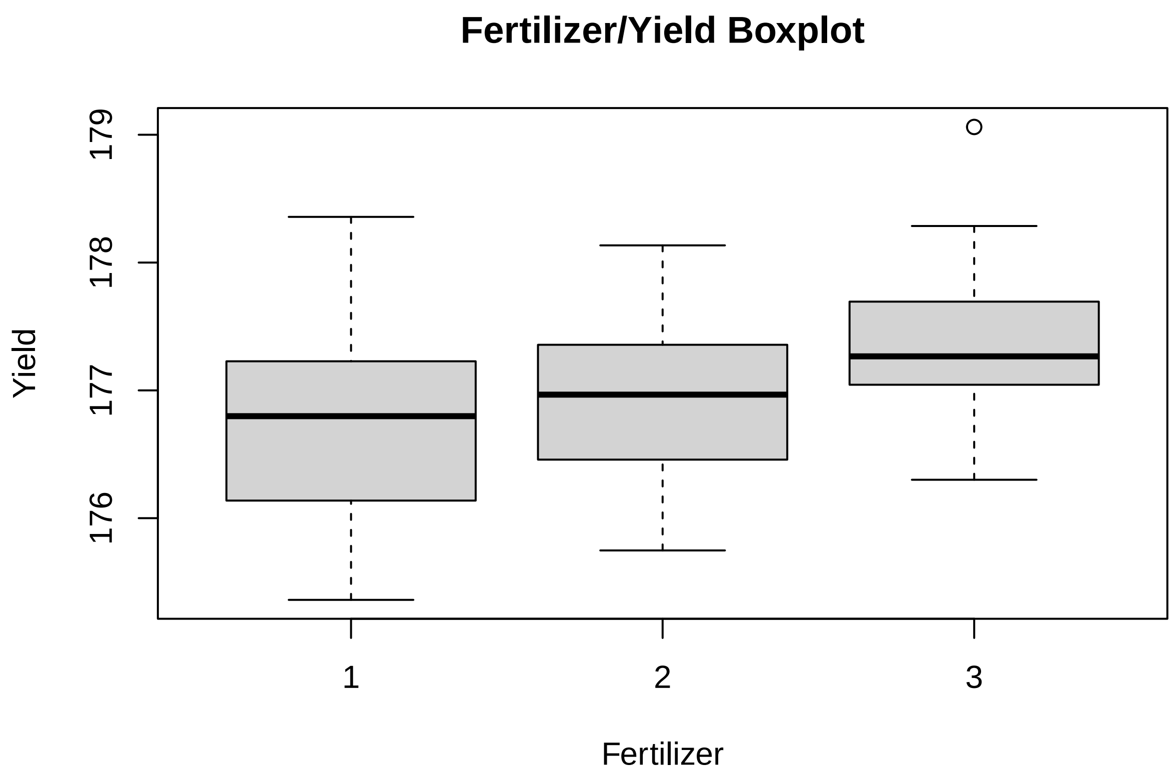
2.1

$H_0: \ \mu_1 = \mu_2 = \mu_3$

H_a : mean yields due to fertilizer are not all equal.

2.2

```
boxplot(xlab = "Fertilizer", ylab = "Yield", CropYield$yield ~ CropYield$fertilizer, beside = TRUE, mai
```



The variance of the yields per fertilizer are varied. From fertilizer 1 through 3, the means seem to increase.

2.3

$$SSB = \sum_{i=1}^{k} n_i(\bar{X}_{i,.} - \bar{X}_{.,.})^2$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{i,j} - \bar{X}_{i,.})^2 = \sum_{n_i - 1} S_i^2$$

$$MSB = \frac{SSB}{k-1}$$

$$MSE = \frac{SSE}{n-k}$$

2.4

```r
summary(aov(yield ~ fertilizer, CropYield))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## fertilizer   2   6.07  3.0340   7.863  7e-04 ***
## Residuals   93  35.89  0.3859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.5

Our testing statistic $F = \frac{MSB}{MSE} = \frac{\frac{SSB}{k-1}}{\frac{SSE}{n-k}}$ is observed as $f = 7.86$. The null distribution is $\mathcal{F}_{k-1,n-k} = \mathcal{F}_{3-1,96-3} = \mathcal{F}_{2,93}$. The p-value is 7×10^{-4}.

2.6

Since $\alpha = 0.05 > 7 \times 10^{-4}$, we must reject H_0 and assume H_a, that is that the mean yields due to fertilizer are not all equal.

2.7

Given means per fertilizer:

```r
tapply(CropYield$yield, CropYield$fertilizer, mean)
```

```
##        1        2        3
## 176.7570 176.9332 177.3562
```

We can check if the mean yield for fertilizer 3 is close to the mean yield across fertilizers 1 and 2, i.e. whether $-\frac{1}{2}\mu_1 - \frac{1}{2}\mu_2 + \mu_3 = 0$.

2.8

```r
contrasts(CropYield$fertilizer) = c(-1/2, -1/2, 1)
go = aov(yield ~ fertilizer, CropYield)
summary(go, split = list(fertilizer = list(`mu3 vs 1/2 mu1 + 1/2 mu2` = 1)))
```

```
##                                        Df Sum Sq Mean Sq F value   Pr(>F)    
## fertilizer                              2   6.07   3.034   7.863 0.000700 ***
##   fertilizer: mu3 vs 1/2 mu1 + 1/2 mu2  1   5.57   5.571  14.439 0.000258 ***
## Residuals                              93  35.89   0.386                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value $0.000258 < \alpha = 0.01$, we must reject H_0 and assume H_a, that fertilizer 3 is different than the mean yield across fertilizers 1 and 2.

2.9

We can do multiple comparison using Fischer's Least Significant Difference.

```r
library(agricolae)
```

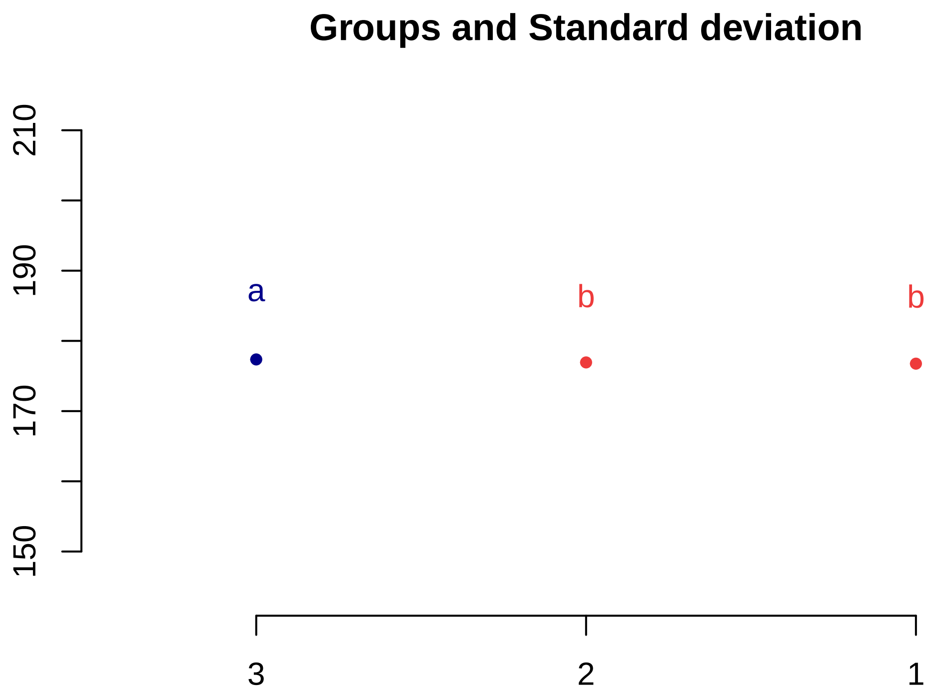
```
## Warning: package 'agricolae' was built under R version 4.4.3
```

```r
comparison = LSD.test(go, "fertilizer", p.adj = "none")
comparison
```

```
## $statistics
##    MSerror Df     Mean        CV  t.value       LSD
##   0.385873 93 177.0155 0.3509223 1.985802 0.3083884
## 
## $parameters
##         test p.ajusted     name.t ntr alpha
##   Fisher-LSD      none fertilizer   3  0.05
## 
## $means
##      yield       std  r        se      LCL      UCL      Min      Max      Q25
## 1 176.7570 0.6849233 32 0.1098113 176.5390 176.9751 175.3608 178.3574 176.1490
## 2 176.9332 0.5740668 32 0.1098113 176.7151 177.1513 175.7475 178.1346 176.4687
## 3 177.3562 0.5991214 32 0.1098113 177.1381 177.5742 176.3005 179.0609 177.0495
##        Q50      Q75
## 1 176.7981 177.2270
## 2 176.9665 177.3547
## 3 177.2661 177.6908
## 
## $comparison
## NULL
## 
## $groups
##      yield groups
## 3 177.3562      a
## 2 176.9332      b
## 1 176.7570      b
## 
## attr(,"class")
## [1] "group"
```

And we can visualize the difference in standard deviation as such:

```r
plot(comparison, variation = "SD")
```


Groups and Standard deviation
150
170
190
210
a
b
b
3
2
1