ISYE 6414 - HW4

Manvitha Kalicheti

gtID: 903838438

Antler Length Data

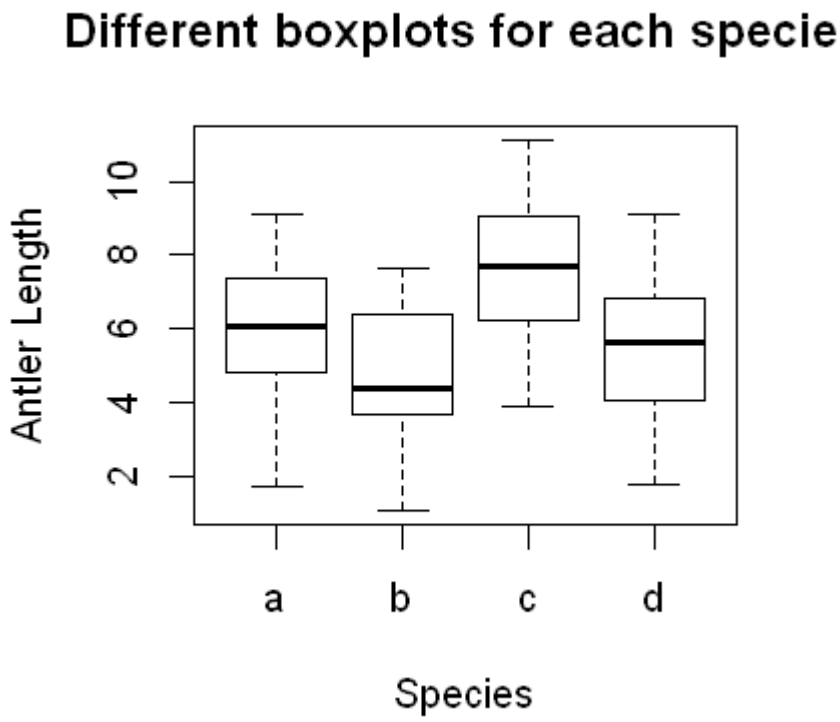
In [1]:

```
data = read.csv('homework04data01.csv')
```

Q1.

In [5]:

```
options(repr.plot.width = 3.5, repr.plot.height = 3.5)
boxplot(Length~Species,
data=data,
main="Different boxplots for each species",
xlab="Species",
ylab="Antler Length")
```



As we can see from the box plots, there is overlap between all 4 of the species, and it is hard to tell if the antler lengths of the four species are different. Although their means are clearly different, their variability is high.

Q2.

In [3]:

```
model = aov(Length ~ Species, data=data)
```

In [4]:

```
summary(model)
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)
Species      3   90.0  30.001   8.121 8.58e-05 ***
Residuals   81  299.2   3.694
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In [10]:

```
model.tables(model, type = 'means')
```

```
Tables of means
Grand mean

6.014706

 Species
         a      b      c      d
     6.063  4.697  7.594  5.445
rep 20.000 15.000 22.000 28.000
```

$H_o : \mu_1 = \mu_2 = \ldots = \mu_k$

$H_a =$ some means are different

$F_0 = \frac{MST_R}{MSE} = \frac{30.001}{3.694} = 8.1215$

From tables, $F_{\alpha,k-1,N-k} = F_{0.05,3,81} = 2.7173$

$F_o >$ critical $F \implies H_o$ can be rejected. Hence, some means are different.

Q3.

From the model summary:

SSE = 299.2

SSTr = 90.0

MSE = 3.694

MSTr = 30.001

Q4.

In [5]:

```
TukeyHSD(model)
```

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

Fit: aov(formula = Length ~ Species, data = data)

$Species
          diff         lwr        upr     p adj
b-a -1.3656667 -3.08777369  0.3564404 0.1683081
c-a  1.5306364 -0.02706805  3.0883408 0.0559481
d-a -0.6176429 -2.09373459  0.8584489 0.6920003
c-b  2.8963030  1.20807683  4.5845292 0.0001303
d-b  0.7480238 -0.86520632  2.3612539 0.6183736
d-c -2.1482792 -3.58469904 -0.7118594 0.0010291
```

c-b and d-c imply a difference in average antler length between species.

Hospital Data

In [6]:

```
data1 = read.csv('homework04Hospital.csv')
```

Q5.

In [8]:

```
model1 = lm(Hours~.,data = data1)
```

In [9]:

```
summary(model1)
```

```
Call:
lm(formula = Hours ~ ., data = data1)

Residuals:
    Min      1Q  Median      3Q     Max 
-677.23 -270.19   60.93  228.32  517.70 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept) 1946.80204  504.18193   3.861  0.00226 **
Xray           0.03858    0.01304   2.958  0.01197 *
BedDays        1.03939    0.06756  15.386 2.91e-09 ***
Length      -413.75780   98.59828  -4.196  0.00124 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 387.2 on 12 degrees of freedom
Multiple R-squared:  0.9961,    Adjusted R-squared:  0.9952 
F-statistic:  1028 on 3 and 12 DF,  p-value: 9.919e-15
```

$\beta_O = 1946.80204$

$\beta_{Xray} = 0.03858$

$\beta_{BedDays} = 1.03939$

$\beta_{Length} = -413.75780$

Q6.

When BedDays and Length are held constant, for every unit increase in Xray, Hours increase by 0.03858.

When Xray and Length are held constant, for every unit increase in BedDays, Hours increase by 1.03939.

When BedDays and Xray are held constant, for every unit increase in Length, Hours decrease by 413.75780.

$B_o = 1946.80204$ is the mean value of Hours when all the other values are 0.

Q7.

For BedDays, p-value = 2.91e-09 << $\alpha = 0.05$ $\implies$ BedDays is statistically significant

$H_0 : \beta_{BedDays} = 0$

$H_a : \beta_{BedDays} \neq 0$

$t - value = 15.386$

critical t = $t_{0.025,12} = 2.179$

$|t - value|$ >critical t $\implies$ H_0 can be rejected $\implies$ BedDays is statistically significant.

Q8.

For Xray, $p - value = 0.01197$

With $\alpha = 0.05$, Xray is statistically significant as $p - value < \alpha$.

But with $\alpha = 0.01$, Xray is not statistically significant as $p - value > \alpha$.

Q9.

For Length, $p - value = 0.00124$

With $\alpha = 0.05$, Length is statistically significant as $p - value < \alpha$.

With $\alpha = 0.10$, Length is statistically significant as $p - value < \alpha$.