Lab #4: Analysis of Variance (ANOVA)

Name

Date of lab session

Lab report

Load data: Load the `InsectSprays` data set into our workspace.

```
data(InsectSprays)
?InsectSprays
chickwts <- chickwts[which(chickwts$feed != "horsebean" & chickwts$feed != "meatmeal"),]
```

Exercises:

Exercise 1: Count and spray are the variables in this data set. There are 72 cases in the sample.

```
dim(InsectSprays)
```

```
## [1] 72  2
```

Exercise 2: We can say that the spray does affect count based on the box plot as there are distinct differences in count depending on the type of insecticide used.
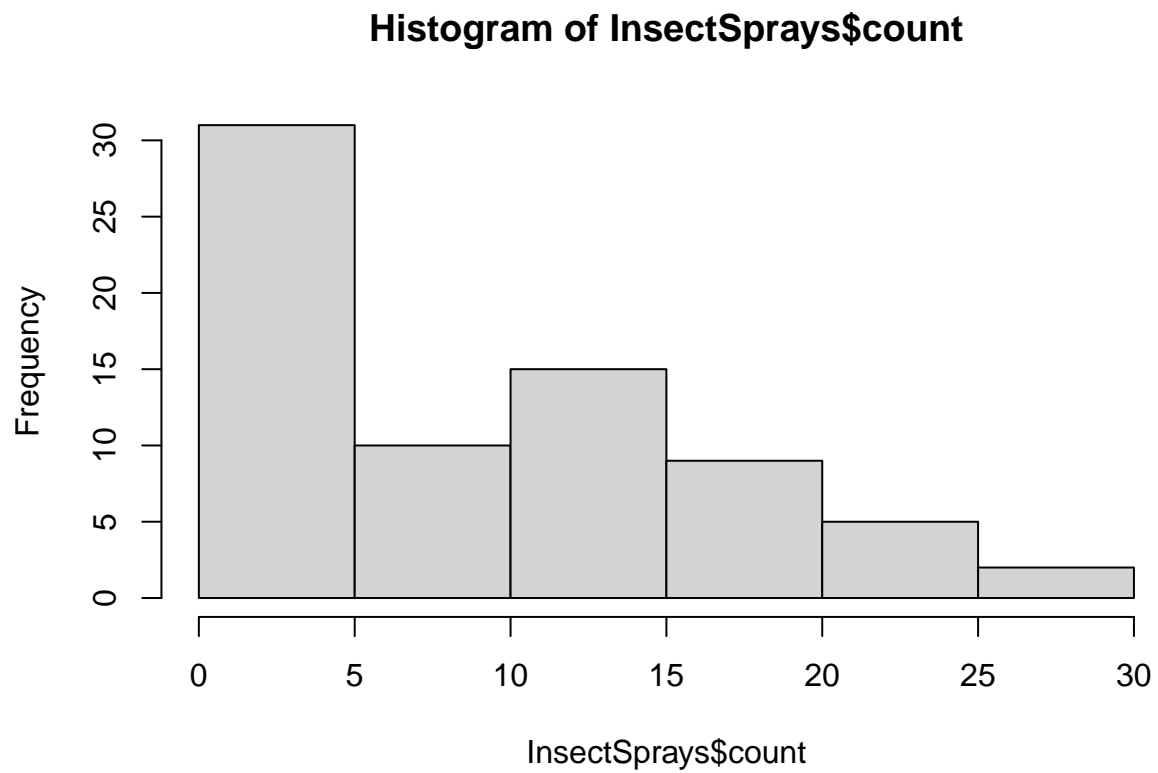
```
summary(InsectSprays)
```

```
##      count          spray
##  Min.   : 0.00    A:12
##  1st Qu.: 3.00    B:12
##  Median : 7.00    C:12
##  Mean   : 9.50    D:12
##  3rd Qu.:14.25    E:12
##  Max.   :26.00    F:12
```
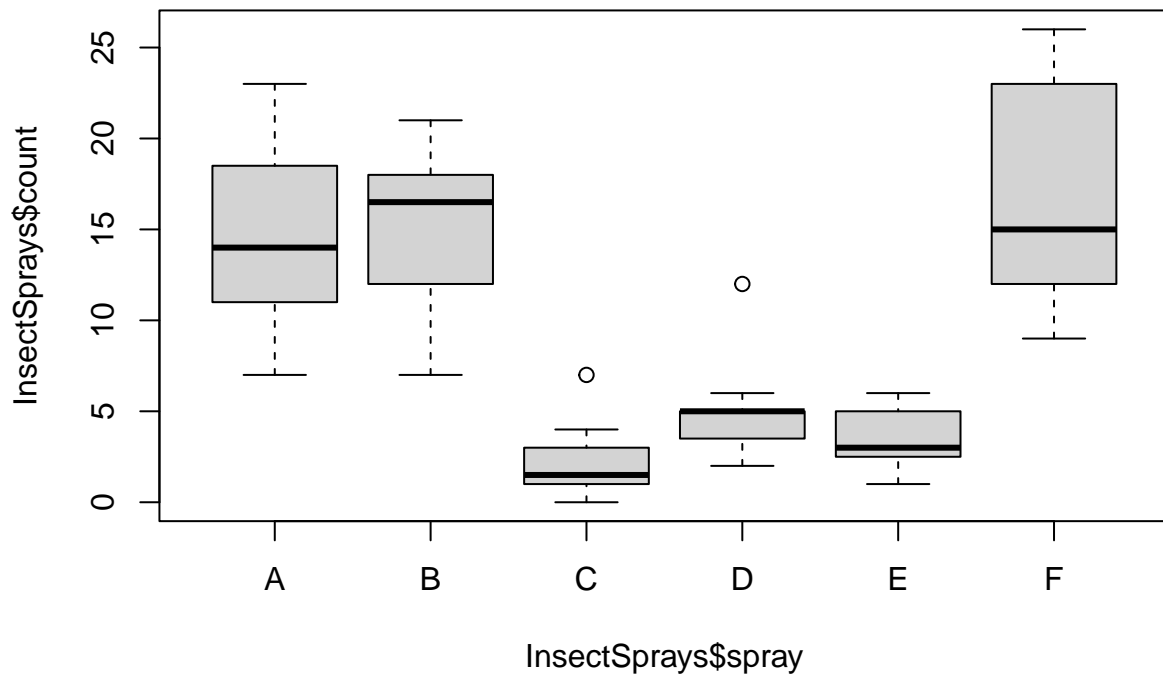
```
dim(InsectSprays)
```

```
## [1] 72  2
```

```r
hist(InsectSprays$count)
```

Histogram of InsectSprays$count


```r
boxplot(InsectSprays$count ~ InsectSprays$spray)
```

Exercise 3: B does not seem to be normally distributed. C does not seem to be normally distributed. D does not seem to be normally distributed. E does not seem to be normally distributed. F does not seem to be normally distributed.
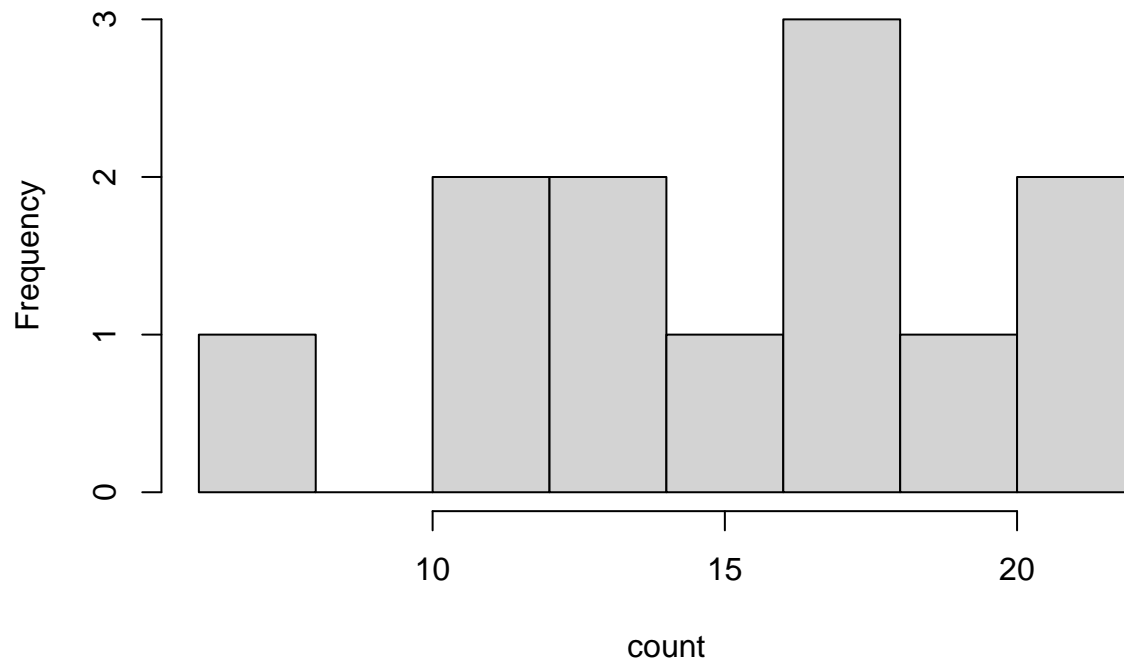
```r
hist(InsectSprays$count[InsectSprays$spray == "A"], breaks=6,
     main ="Histogram of Count for Spray A", xlab="count")
```

Histogram of Count for Spray A
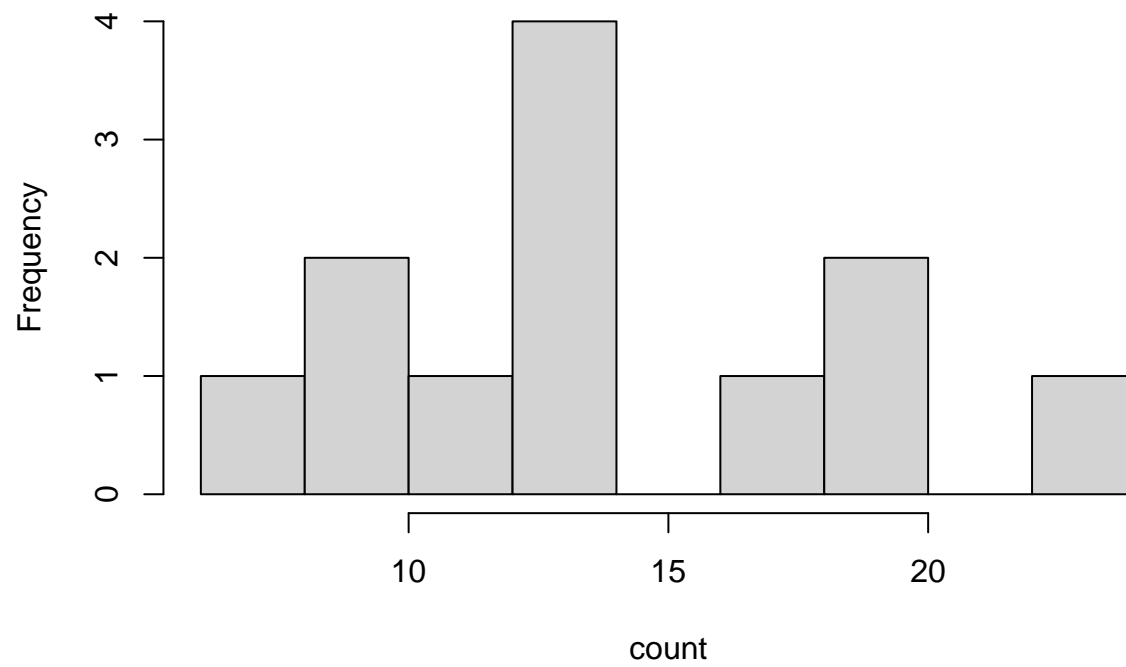

```r
hist(InsectSprays$count[InsectSprays$spray == "B"], breaks=6,
     main ="Histogram of Count for Spray B", xlab="count")
```

Histogram of Count for Spray B


```r
hist(InsectSprays$count[InsectSprays$spray == "C"], breaks=6,
     main ="Histogram of Count for Spray C", xlab="count")
```

Histogram of Count for Spray C


```
hist(InsectSprays$count[InsectSprays$spray == "D"], breaks=6,
     main ="Histogram of Count for Spray D", xlab="count")
```

Histogram of Count for Spray D


```r
hist(InsectSprays$count[InsectSprays$spray == "E"], breaks=6,
     main ="Histogram of Count for Spray E", xlab="count")
```
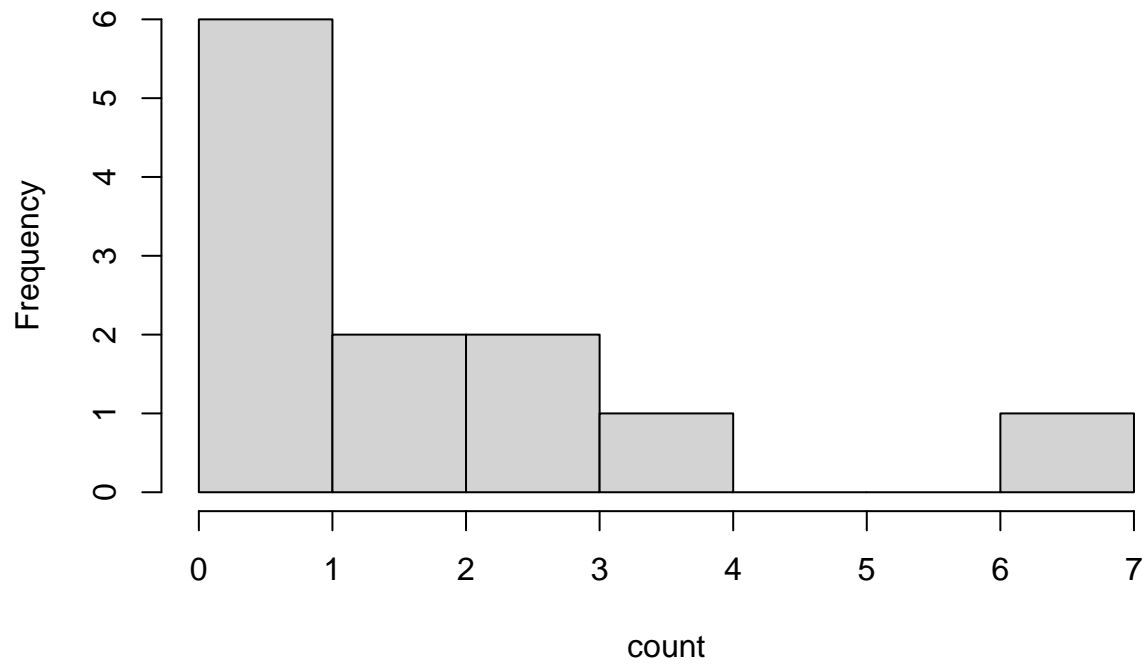
Histogram of Count for Spray E


```r
hist(InsectSprays$count[InsectSprays$spray == "F"], breaks=6,
     main ="Histogram of Count for Spray F", xlab="count")
```

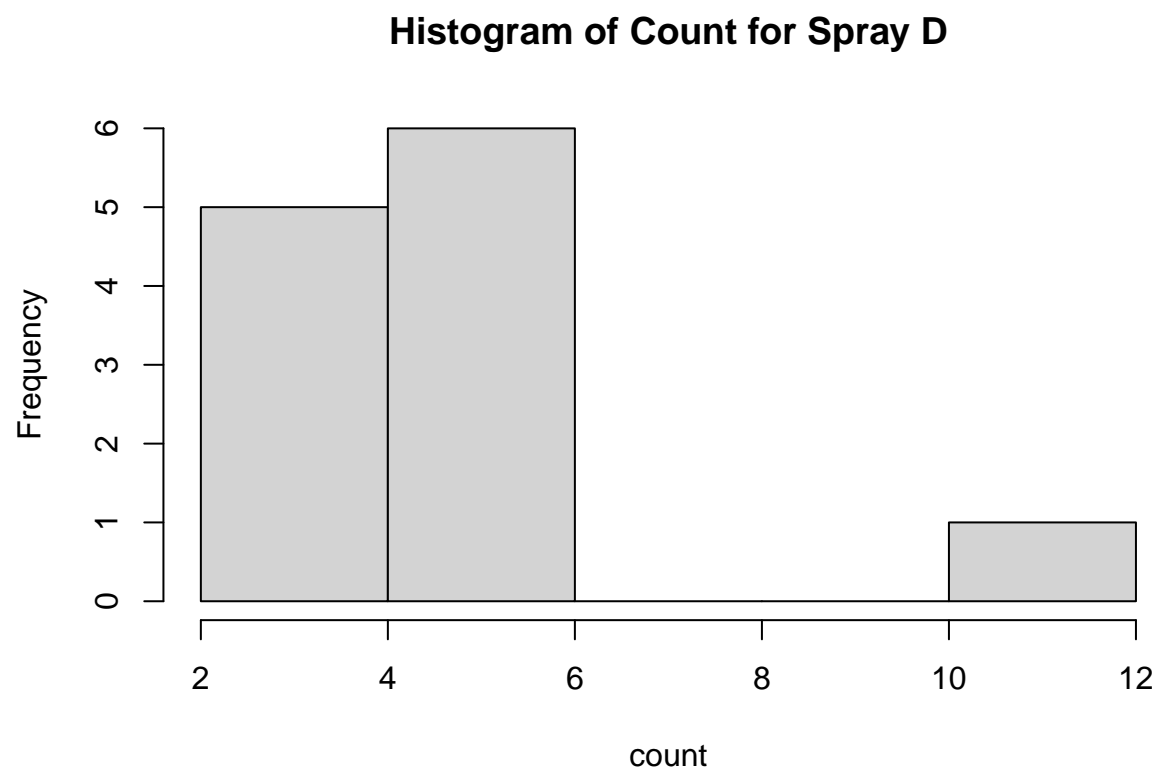# Histogram of Count for Spray F
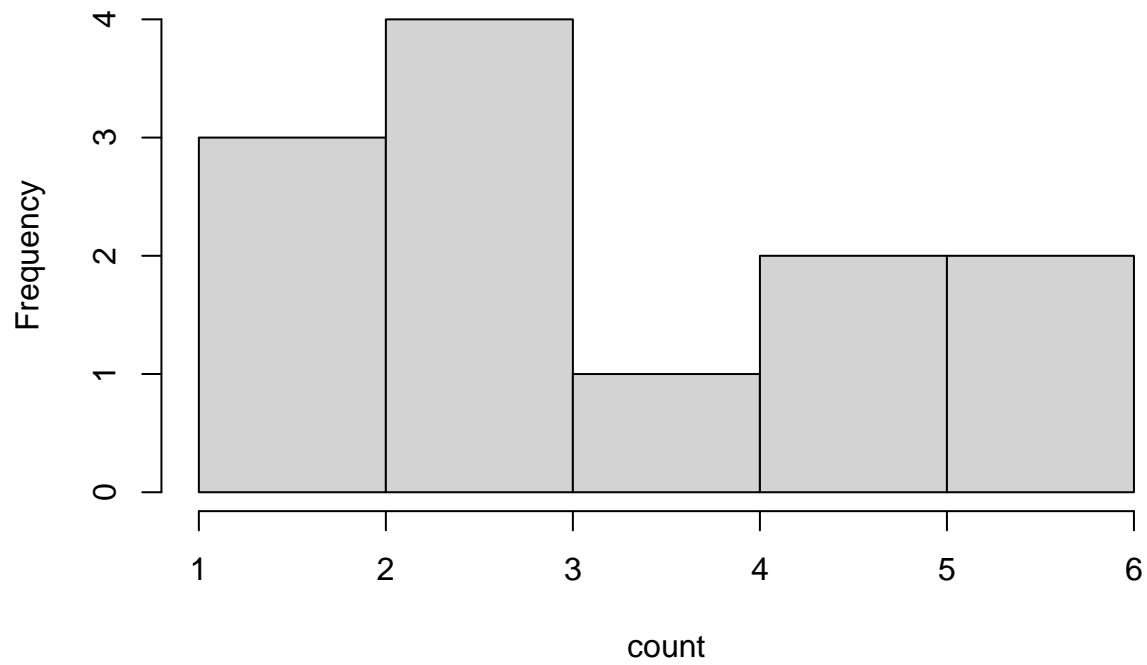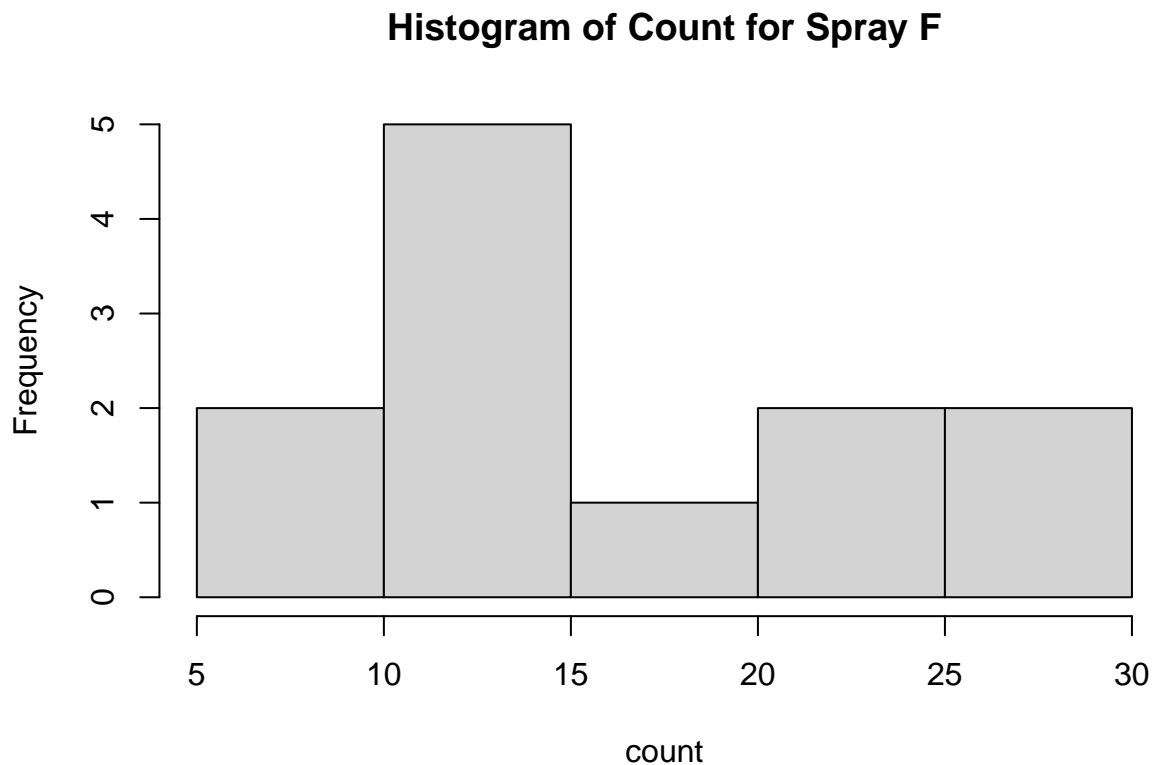

Exercise 4: It is reasonable to assume normality for B. It is not reasonable to assume normality for C. It is not reasonable to assume normality for D. It is reasonable to assume normality for E. It is reasonable to assume normality for F.

```r
shapiro.test(InsectSprays$count[InsectSprays$spray == "B"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  InsectSprays$count[InsectSprays$spray == "B"]
## W = 0.95031, p-value = 0.6415
```

```r
shapiro.test(InsectSprays$count[InsectSprays$spray == "C"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  InsectSprays$count[InsectSprays$spray == "C"]
## W = 0.85907, p-value = 0.04759
```

```r
shapiro.test(InsectSprays$count[InsectSprays$spray == "D"])
```

```
##
```

```
##  Shapiro-Wilk normality test
##
## data:  InsectSprays$count[InsectSprays$spray == "D"]
## W = 0.75063, p-value = 0.002713
```

```r
shapiro.test(InsectSprays$count[InsectSprays$spray == "E"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  InsectSprays$count[InsectSprays$spray == "E"]
## W = 0.92128, p-value = 0.2967
```

```r
shapiro.test(InsectSprays$count[InsectSprays$spray == "F"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  InsectSprays$count[InsectSprays$spray == "F"]
## W = 0.88475, p-value = 0.1009
```

Exercise 5: These are the standard deviations of the sprays in alphabetical order. No, it is not safe to assume that the variances are equal.

```r
afilter <- InsectSprays$spray == 'A'
sd(InsectSprays$count[afilter])
```

```
## [1] 4.719399
```

```r
bfilter <- InsectSprays$spray == 'B'
sd(InsectSprays$count[bfilter])
```

```
## [1] 4.271115
```

```r
cfilter <- InsectSprays$spray == 'C'
sd(InsectSprays$count[cfilter])
```

```
## [1] 1.975225
```

```r
dfilter <- InsectSprays$spray == 'D'
sd(InsectSprays$count[dfilter])
```

```
## [1] 2.503028
```

```r
efilter <- InsectSprays$spray == 'E'
sd(InsectSprays$count[efilter])
```

```
## [1] 1.732051
```

```
ffilter <- InsectSprays$spray == 'F'
sd(InsectSprays$count[ffilter])
```

```
## [1] 6.213378
```

Exercise 6: Null Hypothesis: The mean count is the same across all sprays Alternative Hypothesis: At least one spray has a different mean count from the others.

Exercise 7: We can reject the null hypothesis and conclude that it is unlikely that the mean count is the same across all sprays.

```
summary(aov(InsectSprays$count ~ InsectSprays$spray))
```

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## InsectSprays$spray  5   2669   533.8    34.7 <2e-16 ***
## Residuals          66   1015    15.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
qf(0.95,5,66,lower.tail=FALSE)
```

```
## [1] 0.2259666
```

```
pf(34.7,5,66,lower.tail=FALSE)
```

```
## [1] 3.187508e-17
```

```
pairwise.t.test(x=InsectSprays$count, g=InsectSprays$spray, p.adj="bonf")
```

Exercise 8:

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  InsectSprays$count and InsectSprays$spray
##
##   A       B       C       D       E
## B 1       -       -       -       -
## C 1.1e-09 1.3e-10 -       -       -
## D 1.5e-06 1.8e-07 1       -       -
## E 4.1e-08 4.9e-09 1       1       -
## F 1       1       4.2e-12 6.1e-09 1.6e-10
##
## P value adjustment method: bonferroni
```

A&B A&C Different A&D Different A&E Different A&F B&F C&D C&E * * *

On Your Own

1: Null Hypothesis: The mean weight is the same across all supplements. Alternative Hypothesis: At least one supplement has a different mean weight from the others.

2: Since random sampling is assumed, we can say the cases are all independent from each other. Since all W test statistics are high for the Shapiro-Wilk normality tests are high, and the p-values are higher than the significance level of the Shapiro-Wilk normality tests we can say that all groups are roughly normally distributed. Does not seem to have equal variances according to the box plot.

```r
summary(chickwts)
```

```
##      weight               feed
##  Min.   :141.0   casein   :12
##  1st Qu.:226.8   horsebean: 0
##  Median :271.0   linseed  :12
##  Mean   :278.1   meatmeal : 0
##  3rd Qu.:331.2   soybean  :14
##  Max.   :423.0   sunflower:12
```

```r
dim(chickwts)
```

```
## [1] 50  2
```

```r
shapiro.test(chickwts$weight[chickwts$feed == "linseed"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  chickwts$weight[chickwts$feed == "linseed"]
## W = 0.96931, p-value = 0.9035
```

```r
shapiro.test(chickwts$weight[chickwts$feed == "soybean"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  chickwts$weight[chickwts$feed == "soybean"]
## W = 0.9464, p-value = 0.5064
```

```r
shapiro.test(chickwts$weight[chickwts$feed == "sunflower"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  chickwts$weight[chickwts$feed == "sunflower"]
## W = 0.92809, p-value = 0.3603
```
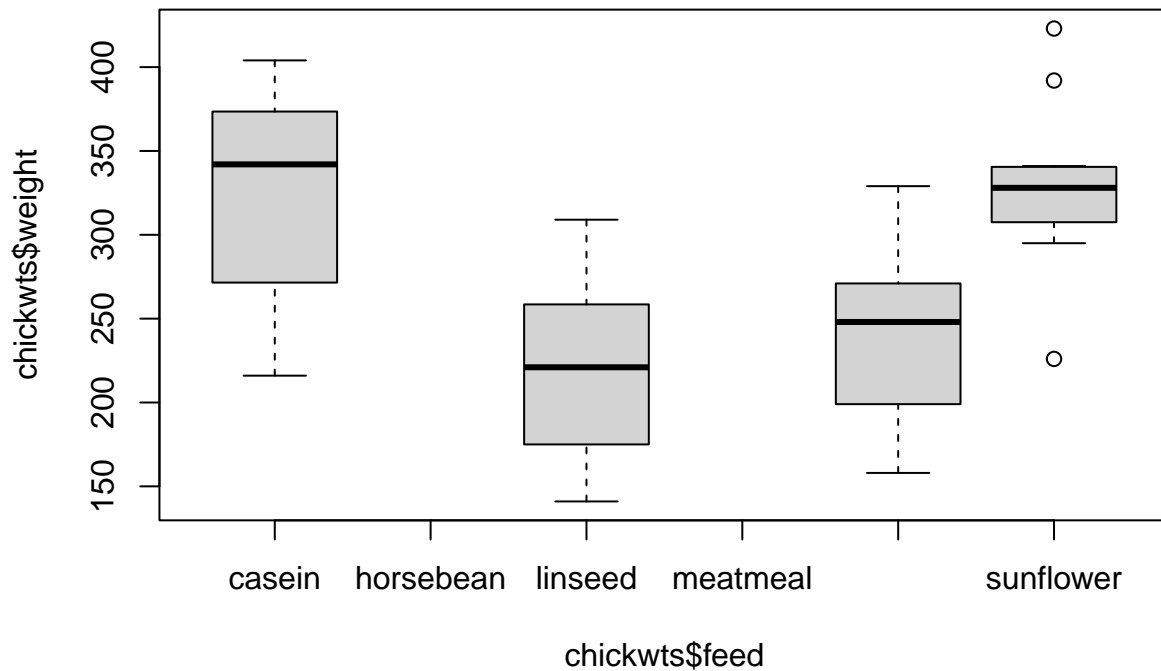
```r
shapiro.test(chickwts$weight[chickwts$feed == "casein"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  chickwts$weight[chickwts$feed == "casein"]
## W = 0.91663, p-value = 0.2592
```

```r
shapiro.test(chickwts$weight[chickwts$feed == "linseed"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  chickwts$weight[chickwts$feed == "linseed"]
## W = 0.96931, p-value = 0.9035
```

```r
boxplot(chickwts$weight ~ chickwts$feed)
```


3: Reject null, at is unlikely that the mean weight is the same across all supplements.

```r
summary(aov(chickwts$weight ~ chickwts$feed))
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## chickwts$feed  3 112125   37375   12.28  5e-06 ***
## Residuals     46 140008    3044
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
pf(12.28,3,46,lower.tail=FALSE)
```

```
## [1] 5.001986e-06
```

Extra Credit

```r
pairwise.t.test(x=chickwts$weight, g=chickwts$feed, p.adj="bonf")
```

1:

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  chickwts$weight and chickwts$feed
##
##           casein  linseed soybean
## linseed   0.00017 -       -
## soybean   0.00533 1.00000 -
## sunflower 1.00000 7.6e-05 0.00254
##
## P value adjustment method: bonferroni
```

Linseed and Casein Soybeen and Casein sunflower and linseed sunflower and soybean