Homework 1

Evan Yacek ety78

This homework uses the `InsectSprays` data set available in R. This data set contains counts of insects in agricultural experimental units (`count`) treated with different insecticides (`spray`). There are six different types of insect sprays tested in the data set (labeled A-F), each with 12 observations.

```
head(InsectSprays)
```

```
##   count spray
## 1    10     A
## 2     7     A
## 3    20     A
## 4    14     A
## 5    14     A
## 6    12     A
```

Question 1: (4 pts) We are interested in testing the effect of insect spray type on the number of insects observed. Since there are six different types of insect sprays in the data set, and therefore six groups of insect counts, we will use an analysis of variance (ANOVA) test. Conduct an ANOVA test and interpret your results in 1-2 sentences. HINT: You will first need to create a linear model object using the `lm()` function before you can use the `anova()` function.

```
# your R code goes here
fit <- lm(count~spray, data = InsectSprays)
summary(fit)
```

```
##
## Call:
## lm(formula = count ~ spray, data = InsectSprays)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.333 -1.958 -0.500  1.667  9.333
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.5000     1.1322  12.807  < 2e-16 ***
## sprayB        0.8333     1.6011   0.520    0.604
## sprayC      -12.4167     1.6011  -7.755 7.27e-11 ***
## sprayD       -9.5833     1.6011  -5.985 9.82e-08 ***
## sprayE      -11.0000     1.6011  -6.870 2.75e-09 ***
## sprayF        2.1667     1.6011   1.353    0.181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.922 on 66 degrees of freedom
## Multiple R-squared:  0.7244, Adjusted R-squared:  0.7036
## F-statistic:  34.7 on 5 and 66 DF,  p-value: < 2.2e-16
```

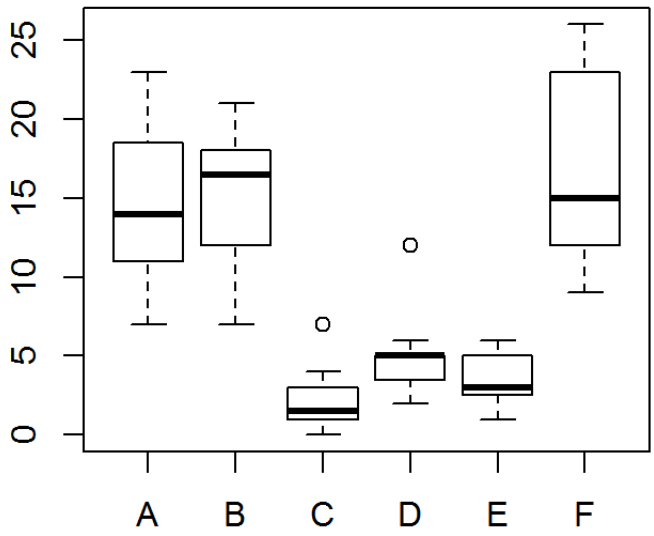
```
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: count
##           Df Sum Sq Mean Sq F value    Pr(>F)
## spray      5 2668.8  533.77  34.702 < 2.2e-16 ***
## Residuals 66 1015.2   15.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Spray C,E, and F are the most effective sprays to limit insect count as shown by the negative values on the estimate on the summary of the linear model. Furthermore, the anova shows that the sprays did impact the the mean insect count.

Question 2: (3 pts) Create a boxplot of the insect count data, separated by spray type. Based on this plot, is the mean insect count of Spray D the same or different from that of Spray F? Explain your answer.

```
insects <- unstack(InsectSprays)
boxplot(insects)
```



The mean insect count of spray D is significantly lower than spray F. Box plot F shows that the mean insect count hovers around 15 while Spray D's insectc count was around 3.

Question 3: (3 pts) Use a t test to determine if the mean insect count of Spray D is the same or different from that of Spray F. Interperet and explain your results in 1-2 sentences.

```
t.test(insects$D, insects$F)
```

```
##
##  Welch Two Sample t-test
##
## data:  insects$D and insects$F
## t = -6.0764, df = 14.479, p-value = 2.47e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -15.884588  -7.615412
## sample estimates:
## mean of x mean of y
##  4.916667 16.666667
```

The significantly low p-value rejects the null hypothesis that both D and F have the same mean insect count. Thus their means must not be equal.