

# Assignment 7

Simon Hans Edasi

2023-06-02

NOTE: Throughout this and subsequent analysis sections of the assignments, the spaces for your answers will look like the space below.

You may increase the space by using the Enter key on your keyboard. Be sure to prefix any written answers to questions with the “#” command to avoid R errors. Your reply should have each line prefaced by a “#”. Please remember to press the ENTER key when you wish to go to the next (or a new) line. Otherwise, R will just let you type on the same line forever!!

Please try to keep all of your typing confined to what is easily visible in RStudio without having to scroll to the right so that your work is easily readable. Notice how I have typed my comments and code. I have tried to make it readable without scrolling to the right.

NOTE:

Please read Chapters 5.4 and 5.6 in BC&P while doing the “ANOVA Tutorial Key.r” as you read the chapters. Then do this assignment.

Two excellent general references are:

Jerrold H. Zar. 2010 (5th edition). Biostatistical Analysis. Prentice Hall. (This is often the text for Q SCI 482.)

Gerald van Belle, L. D. Fisher, P. J. Heagerty, and T Lumley. 2004 (2nd edition). Biostatistics: A Methodology for the Health Sciences. Wiley.

Both of these books provide a broad coverage of statistics including ANOVA and linear regression.

Introduction – the Experiment:

A completely randomized experiment was conducted to evaluate the effects of different insect sprays on insect prevalence in agricultural plots. You may wish to review the class notes for Chapter 10 regarding the definition of a completely randomized experimental design.

Agricultural plots were randomly selected and assigned at random to receive six different insect spray treatments, one of which is the control (no treatment). There are 72 plots in the experiment. Each spray was assigned to 12 plots ( $6 \times 12 = 60$ ) and the control was assigned to 12 plots. In this case the treatment actually consists of treating something. The plants are sprayed with an insecticide or no insecticide. The control (no insecticide spray used) is also considered to be a treatment. You should also notice that the experiment is well replicated (there are 12 replicates for each treatment). After the treatment, the average number of insects found on the plants is recorded. We have 12 random samples from 6 populations

In order to maintain the selected alpha level while comparing the means of more than two groups, an analysis of variance (ANOVA) is used rather than multiple pair-wise t-tests (for details, see Zar Chapter 10).

We are going to use a one-way ANOVA to compare the means of the insect counts in the plots. The ANOVA is one-way because we are analyzing the effect of only one factor (the insecticide) on the abundance of insects on the plants after the treatment.

The experimental data collected is found in the data set “InsectSprays.csv”.

The data set consists of three columns: “number” identifies the plot number; “rep” represents the replicate number; “count” records the average number of insects; and “spray” identifies the treatment ( A through F).

In the space below, please clear the R memory; load the libraries “ggplot2”, “dplyr”, “ggfortify”, and “readr”, then load the dataset, and assign it to an object named “Insect\_Spray”.

```
#Clear the memory
rm(list = ls())

##
# Load libraries
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggfortify)
library(readr)
```

Load the InsectSprays.csv data into R:

```
Insect_Spray <- read.csv("InsectSprays.csv")
```

Now that we have the data in R, let’s take a quick look at it. We will want to find out such things as:

[1] What does the dataset look like? [2] what type of data is it? [3] What are the column names in the dataset? [4] How many rows and columns are in the dataset?

Feel free to look at anything else you would like to know about the dataset.

Please enter your code in the space below.

```
# View(Insect_Spray)
str(Insect_Spray)

## 'data.frame':   72 obs. of  4 variables:
## $ number: int  1 2 3 4 5 6 7 8 9 10 ...
## $ rep   : int  1 2 3 4 5 6 7 8 9 10 ...
## $ count : num  10 7.1 20.2 14.1 14.1 12.3 10 23.1 17.2 20 ...
## $ spray : chr  "A" "A" "A" "A" ...
```

```
colnames(Insect_Spray)
```

```
## [1] "number" "rep" "count" "spray"
```

```
nrow(Insect_Spray)
```

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

```
ncol(Insect_Spray)
```

```
## [1] 4
```

## Part 1: Data Exploration

As always, the first step in the analysis will be to explore the data looking for the structure of the dataset (what is in it), outliers, and if it appears to meet the assumptions of the hypothesis test we will apply in analyzing it. Plotting the data is critical in the exploration as are summary statistics.

From the ANOVA Tutorial, here is one way to do this. Run the script.

```
by(Insect_Spray, Insect_Spray$spray, summary)
```

```
## Insect_Spray$spray: A
##      number      rep      count      spray
##  Min.   : 1.00   Min.   : 1.00   Min.   : 7.10   Length:12
## 1st Qu.: 3.75   1st Qu.: 3.75   1st Qu.:11.72   Class :character
## Median : 6.50   Median : 6.50   Median :14.10   Mode  :character
## Mean   : 6.50   Mean   : 6.50   Mean   :14.64
## 3rd Qu.: 9.25   3rd Qu.: 9.25   3rd Qu.:17.90
## Max.   :12.00   Max.   :12.00   Max.   :23.10
## -----
## Insect_Spray$spray: B
##      number      rep      count      spray
##  Min.   :13.00   Min.   : 1.00   Min.   : 7.10   Length:12
## 1st Qu.:15.75   1st Qu.: 3.75   1st Qu.:12.62   Class :character
## Median :18.50   Median : 6.50   Median :16.55   Mode  :character
## Mean   :18.50   Mean   : 6.50   Mean   :15.50
## 3rd Qu.:21.25   3rd Qu.: 9.25   3rd Qu.:17.75
## Max.   :24.00   Max.   :12.00   Max.   :21.30
## -----
## Insect_Spray$spray: C
##      number      rep      count      spray
##  Min.   :25.00   Min.   : 1.00   Min.   :0.00   Length:12
## 1st Qu.:27.75   1st Qu.: 3.75   1st Qu.:1.00   Class :character
## Median :30.50   Median : 6.50   Median :1.55   Mode  :character
## Mean   :30.50   Mean   : 6.50   Mean   :2.10
## 3rd Qu.:33.25   3rd Qu.: 9.25   3rd Qu.:3.00
## Max.   :36.00   Max.   :12.00   Max.   :7.10
## -----
## Insect_Spray$spray: D
##      number      rep      count      spray
##  Min.   :37.00   Min.   : 1.00   Min.   : 2.000   Length:12
## 1st Qu.:39.75   1st Qu.: 3.75   1st Qu.: 3.775   Class :character
## Median :42.50   Median : 6.50   Median : 5.000   Mode  :character
```

```
## Mean :42.50 Mean : 6.50 Mean : 4.917
## 3rd Qu.:45.25 3rd Qu.: 9.25 3rd Qu.: 5.125
## Max. :48.00 Max. :12.00 Max. :12.400
## -----
## Insect_Spray$spray: E
##      number      rep      count      spray
## Min. :49.00 Min. : 1.00 Min. :0.900 Length:12
## 1st Qu.:51.75 1st Qu.: 3.75 1st Qu.:2.750 Class :character
## Median :54.50 Median : 6.50 Median :3.250 Mode :character
## Mean :54.50 Mean : 6.50 Mean :3.600
## 3rd Qu.:57.25 3rd Qu.: 9.25 3rd Qu.:5.175
## Max. :60.00 Max. :12.00 Max. :6.100
## -----
## Insect_Spray$spray: F
##      number      rep      count      spray
## Min. :61.00 Min. : 1.00 Min. : 9.10 Length:12
## 1st Qu.:63.75 1st Qu.: 3.75 1st Qu.:13.15 Class :character
## Median :66.50 Median : 6.50 Median :15.55 Mode :character
## Mean :66.50 Mean : 6.50 Mean :17.20
## 3rd Qu.:69.25 3rd Qu.: 9.25 3rd Qu.:23.27
## Max. :72.00 Max. :12.00 Max. :26.70
```

Please try to create another way to do this using “dplyr”.

In the space below, create an R script that produces for each treatment a summary of the data including means, medians, IQR’s and anything else you may wish to display. You may wish to review Assignment 5 for ideas.

```
Insect_Spray %>%
  group_by(spray) %>%
  summarise(
    total_count = sum(count),
    count_IQR = IQR(count),
    count_min = min(count),
    count_mean = mean(count),
    count_max = max(count),
  )

## # A tibble: 6 x 6
##   spray total_count count_IQR count_min count_mean count_max
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 A         176.         6.17         7.1         14.6         23.1
## 2 B         186         5.12         7.1         15.5         21.3
## 3 C          25.2          2          0          2.1          7.1
## 4 D          59         1.35          2          4.92         12.4
## 5 E         43.2         2.42          0.9          3.6          6.1
## 6 F        206.        10.1          9.1         17.2         26.7
```

Graph the data in whatever way(s) you find most relevant. Make sure to do a boxplot.

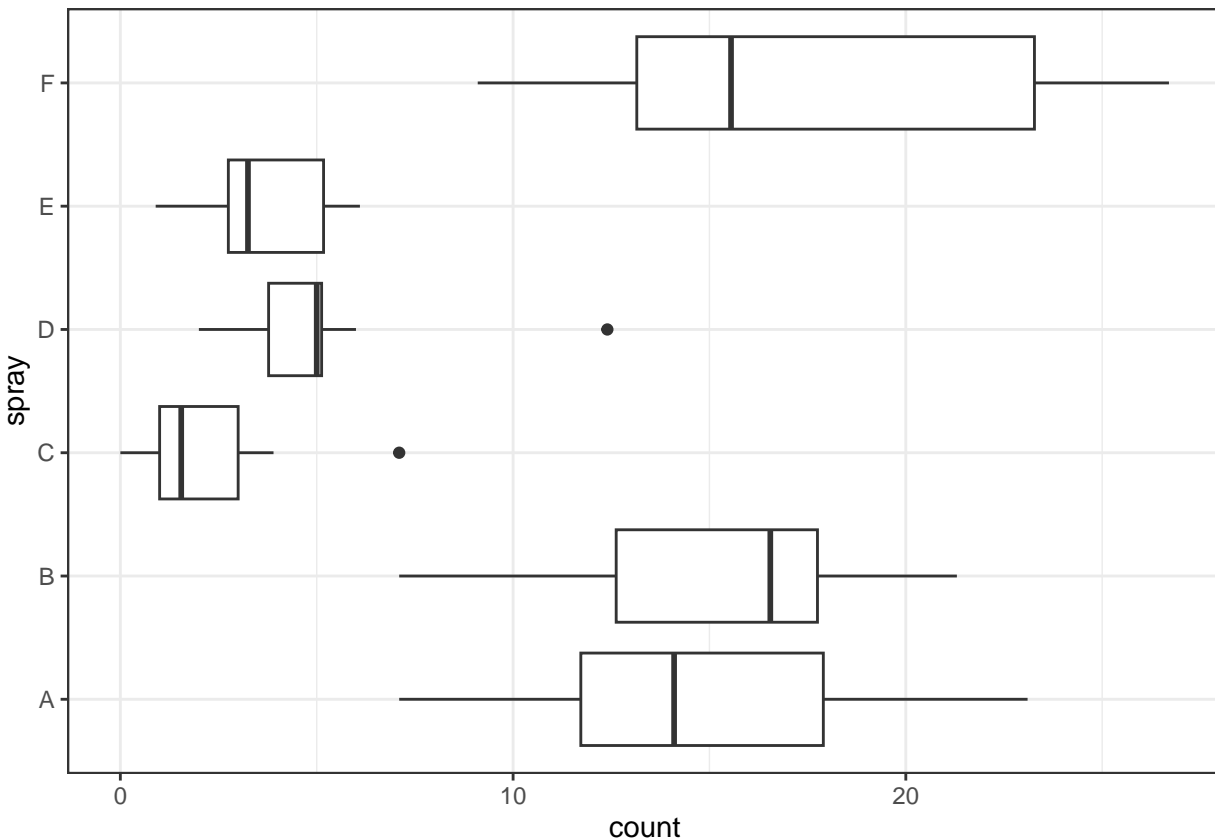
For this boxplot you do not have to flip the axes, but you might try it both ways just for practice with “ggplot” and its parameters.

You might label the x-axis “Spray Treatment” and the y-axis “Number of Insects”, but the choice of axis labels is ultimately yours.

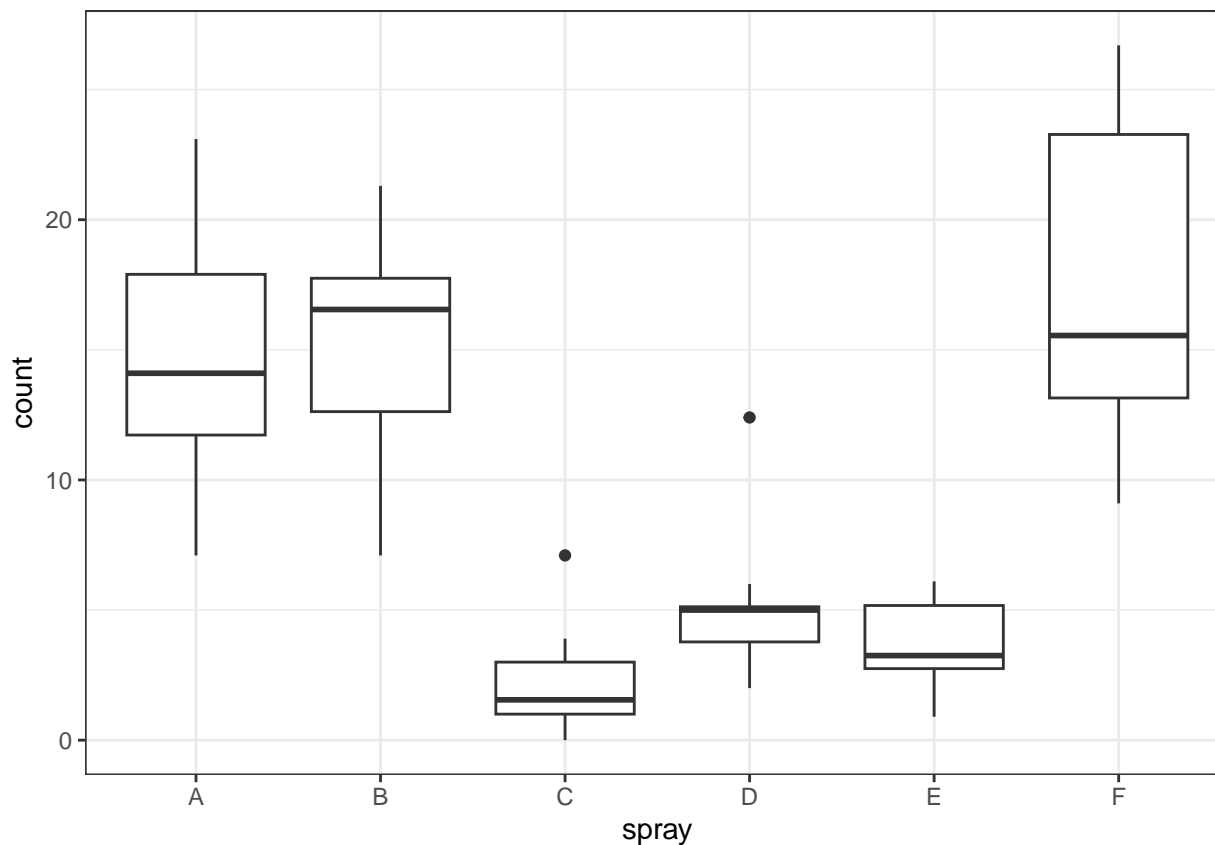
Then, going back to previous assignments and tutorials, create some code that will incorporate an identification of potential outliers into the boxplot. If your code identifies any outliers, do not forget what you have learned to do when you encounter them, but we will accept them as correct for the purposes of this assignment and continue on with the analysis of the experiment.

Please put your script in the space below.

```
# Flip the axes  
ggplot(Insect_Spray, aes(x = spray, y = count)) +  
  geom_boxplot() +  
  theme_bw() +  
  coord_flip()
```

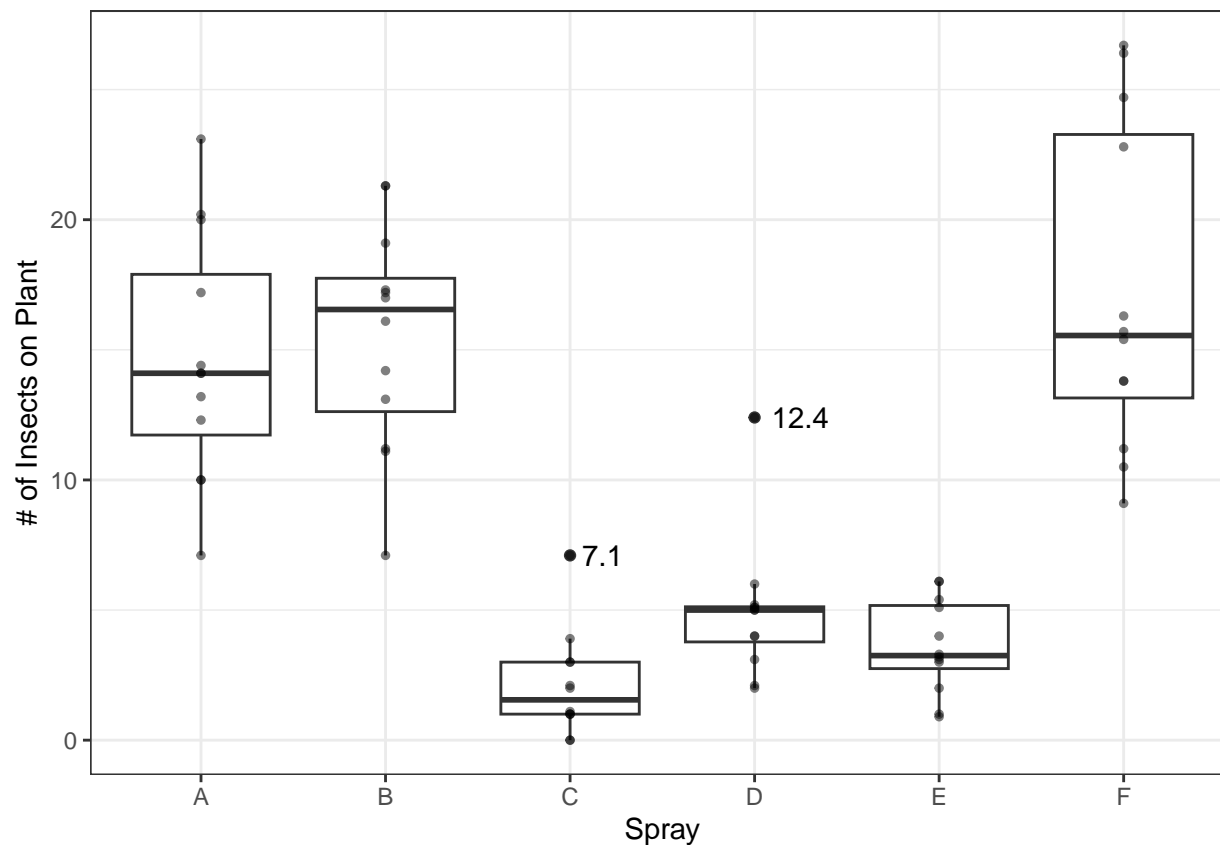


```
# Don't flip the axes  
ggplot(Insect_Spray, aes(x = spray, y = count)) +  
  geom_boxplot() +  
  theme_bw()
```



```
# coord_flip()

# Incorporate an identification of any outliers in the boxplot.
is_outlier <- function(x) {
  return(x < quantile(x,0.25) - 1.5 * IQR(x) |
         x > quantile(x, 0.75) + 1.5 * IQR(x))
}
Insect_Spray %>%
  group_by(spray) %>%
  mutate(outlier = ifelse(is_outlier(count), count, as.numeric(NA))) %>%
  ggplot(., aes(x = factor(spray), y = count)) +
  geom_boxplot() +
  geom_text(aes(label = outlier), na.rm = TRUE, hjust = -0.3)+
  geom_point(size = 1, colour = 'black', alpha = 0.5) +
  xlab('Spray') +
  ylab("# of Insects on Plant") +
  theme_bw()
```



```
# coord_flip()
```

Questions:

[1] Based strictly on your plots, do you think that the sprays have a different effect on the insect count? If so, which ones? Which spray treatment appears to be the most effective in reducing the average insect abundance?

[2] Based strictly on your plots which treatment do you think is the control (no insecticide applied)? Why do you conclude this?

[3] Does your outlier analysis identify any outliers? If so, where?

Please place your answers in the space below.

Spray C, D, and E seem to have a significant effect on the insects on plants.

Spray A and B have a reduced variance, but doesn't look significant

I think spray F is the control. It has the highest number of insects and the highest variance of data.

yes, in sprays C and D there are outliers

The ANOVA assumes:

1. The data consists of independent random samples;
2. There should be no significant outliers;
3. The data should be approximately normally distributed; and
4. The treatment variances are equal.

Questions:

[4] Based on the plots, which variable is the dependent variable and which is the independent variable? Justify your answers.

[5] Is assumptions 1 met? Why or why not?

[6] Based on the plots, explicitly state which of the remaining assumptions are met and which are not. Does your answer to this cause any concern regarding using the ANOVA? If so, why or why not?

Please place your answers to these questions in the space below.



The dependent variable is the spray A - F, and the independent variable is the count of insects found on plants after spraying

No, the assumption is not met because we have found outliers in spray

C and D

The data do appear normally distributed from the boxplots

The variances could potentially be unequal, as the data are spread more

in boxplots with higher values of counts of insects.

These could cause concern for the ANOVA test as the variances are unequal

## Part 2: ANOVA Test

We are going to use a one-way ANOVA to test whether the sprays have the same or different effects as insecticides.

Assume we select alpha to equal 0.05.

We will apply the five step procedure for hypothesis testing in our analysis. As a reminder, those steps, in slightly modified form, are:

Step 1: State your null and alternative hypothesis.

Step 2: Verify that all assumptions necessary for the analysis have been met or that we have stated reasons why we can use the procedure even if they have not been met. If met, calculate the appropriate test statistic. We have done some of this and we will do some more. We will use R to calculate the test statistic.

Step 3: Find the critical value using the appropriate degrees of freedom for the F distribution (in the case of ANOVA). Hint: for the F critical value, with  $k$  = number of treatments and  $n$  = number of observations, the numerator degrees of freedom is  $k-1$  and  $k(n-1)$  is the denominator degrees of freedom. Remember, the F statistic is in the right tail.

Step 4: Report the F statistic and compare it to critical value. State the decision rule.

Step 5: State your conclusion in the context of the problem.

Question (Step 1):

[7] state the null and alternative hypotheses.

Please place your answer in the space below.

Answer (Step 1):

**H<sub>0</sub>:** there is no difference in mean insect count between spray treatments

**H<sub>a</sub>:** There is at least one difference in means between spray treatments

Step 2: We have done some of this and we will do some more below. We will use R to calculate the test statistic.

Define the linear model and assign it to an object named “model\_spray”, then display the results.

Once we have done this, we can create and display some diagnostic plots to further evaluate the validity of the assumptions for the ANOVA.

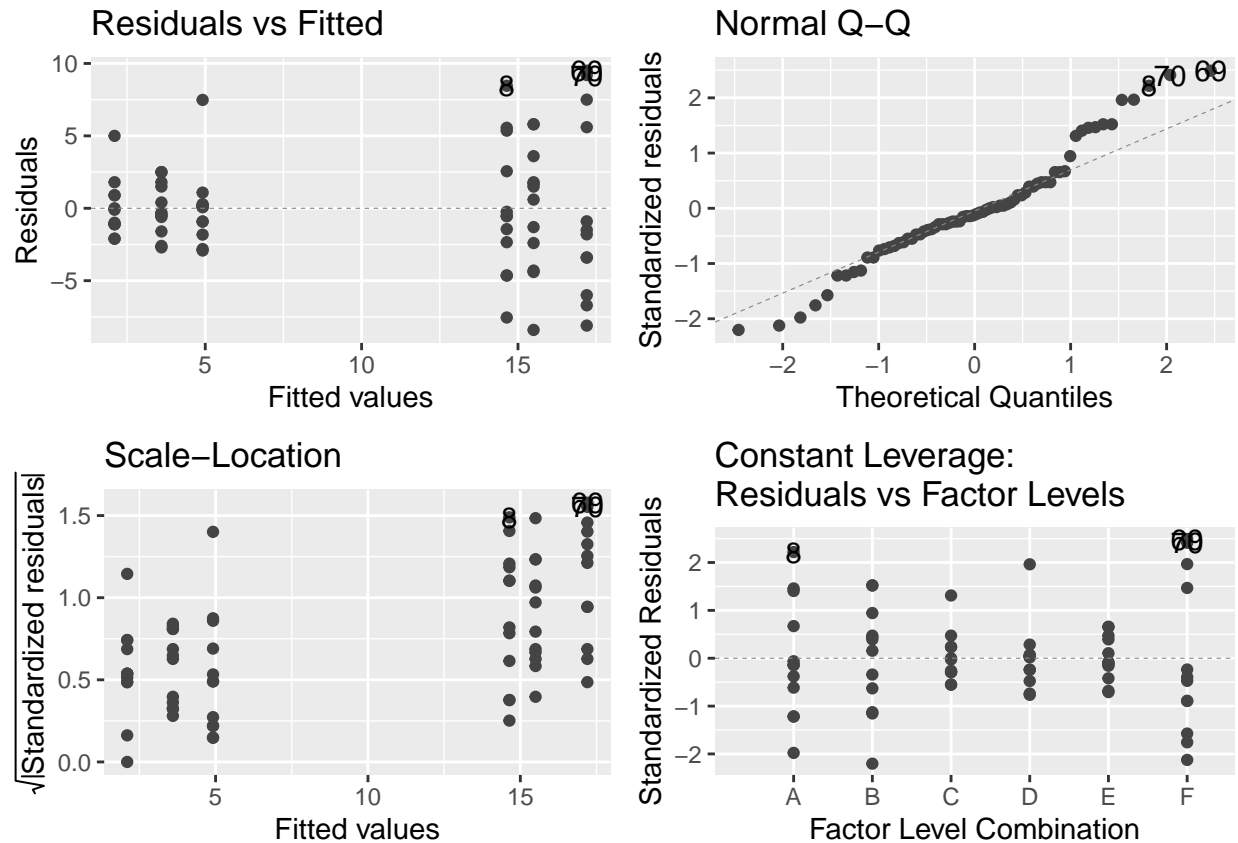
```
model_spray <- lm(count ~ spray, data = Insect_Spray)
model_spray
```

```
##
## Call:
## lm(formula = count ~ spray, data = Insect_Spray)
##
## Coefficients:
## (Intercept)      sprayB      sprayC      sprayD      sprayE      sprayF
##    14.6417      0.8583     -12.5417     -9.7250     -11.0417      2.5583
```

```
##
# Plot four diagnostic plots that
# examine the residuals.
##
```

```
autoplot(model_spray, smooth.colour = NA)
```

```
## Warning: Removed 72 rows containing missing values (`geom_line()`).
## Removed 72 rows containing missing values (`geom_line()`).
```



Now examine these plots. All of these plots are based on analyzing residuals-errors.

Based on what we learned about these four plots in the ANOVA Tutorial, an examination of them shows the some assumptions for the ANOVA are not comfortably met.

Note: In the wild (the real world) we may opt to use a nonparametric procedure (one the does not assume anything about the distribution of the underlying random variables) to avoid these issues. An example is the Kruskal-Wallis test, which is a nonparametric version of the ANOVA. All of these considerations are beyond the scope of our class.

We will rely upon the robustness of the ANOVA, especially since the sample sizes are equal, to assume that we may validly use the ANOVA to analyze the experimental data. Welcome to the real world!

Step 3: Find the critical value using the appropriate degrees of freedom are for the F distribution (in the case of ANOVA). Hint: for the F critical value, with  $k$  = number of treatments and  $n$  = number of observations, the numerator degrees of freedom is  $k-1$  and  $k(n-1)$  is the denominator degrees of freedom. Remember, the F statistic is in the right tail.

Calculate the number of observations (sample size, =  $n$ ) for each treatment, then print the result.

```
Num_obs <- Insect_Spray %>%
  group_by (rep) %>%
  summarise(n())%>%
  nrow()
```

In the space below, please display the result of running the above script. Make sure to manually check that these are correct.

```
Num_obs
```

```
## [1] 12
```

Calculate the number of treatments (= k), then print the result.

```
Num_treatments <- Insect_Spray %>%  
  group_by(spray) %>%  
  summarise(n()) %>%  
  nrow()
```

In the space below, please display the result of running the above script. Make sure to manually check that these are correct.

```
Num_treatments
```

```
## [1] 6
```

We know that the degrees of freedom (DOF) are: Numerator DOF = Number of treatments - 1 Denominator DOF = Number of treatments \* (Number of observations - 1)

```
Num_DOF <- Num_treatments-1  
Den_DOF <- Num_treatments*((Num_obs)-1)
```

In the space below, please display the result of running this script. Make sure to manually check that these are correct.

```
print(Num_DOF)
```

```
## [1] 5
```

```
print(Den_DOF)
```

```
## [1] 66
```

In the space below, write R code to declare the alpha value to be 0.05; assigning this value to the object "Alpha\_value".

```
Alpha_value = 0.05
```

Calculate the critical value of F with the selected alpha and the calculated DOF

```
Critical_F_Value <- qf(1-Alpha_value,  
                      Num_DOF,Den_DOF)
```

Display the critical value in the space below

```
Critical_F_Value
```

```
## [1] 2.353809
```

```
# Create ANOVA table and print the results.
```

```
print(ANOVA_Table <- anova(model_spray))
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: count
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## spray      5 2785.8  557.16  35.128 < 2.2e-16 ***
```

```
## Residuals 66 1046.8   15.86
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Display the ANOVA table

```
ANOVA_Table
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: count
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## spray      5 2785.8  557.16  35.128 < 2.2e-16 ***
```

```
## Residuals 66 1046.8   15.86
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Step 4: Report the F statistic and compare it to critical value. State the decision rule and its conclusion.

In the space below, please complete this step.

**The F statistic is higher than the critical value, so we reject the null hypothesis that the means are statistically equal.**

```
# Find the treatment means.
```

```
sumDat <- Insect_Spray %>%
```

```
  group_by(spray) %>%
```

```
  summarise(meanAbundance = mean(count))
```

Display the results

```
sumDat
```

```
## # A tibble: 6 x 2
```

```
##   spray meanAbundance
```

```
##   <chr>          <dbl>
```

```
## 1 A             14.6
```

```
## 2 B             15.5
```

```
## 3 C              2.1
```

```
## 4 D             4.92
```

## 5 E	3.6
## 6 F	17.2

Step 5: State your conclusion in the context of the problem.

Please complete this step in the space below.

**We can conclude that there is a statistically significant difference of mean**

**number of insects when sprayed with different pesticides.**

Question 8: Are you able to conclude which sprays are most effective based on the ANOVA output? Explain.

Please place your answer to this question in the space below.

**Based on the ANOVA output, no. The ANOVA output only gives us a test statistic,**

**we need to do more to determine which spray has a different mean.**

Assume that you do not know about Tukey's HSD test and that you do not insist that the alpha value remains as selected. Suggest a procedure, using your arsenal of statistical knowledge and tests, which would allow you to come to the conclusion about which spray is the most effective. You can perform as many tests as you like. Briefly explain what you would expect the outcome of these tests to be.

**I would run a t-test on each spray against the assumed control F. I expect**

**that these tests would show a different mean for each spray.**

Now use Tukey's HSD test to determine which treatment is the most effective as an insecticide. We know that Tukey's procedure will test the combination of all pairs of treatments.

Given what you have learned about counting in the probability, how many such combinations do you expect to be compared? No fair looking ahead in this assignment to see the answer!! Make use of the hint above!

In the space below, please enter you answer.

ANSWER HERE

```

# Add these new R libraries and the
# Tukey HSD test analysis can be run
# using the following functions.
#
##

library(agricolae)
library(rstatix)

##
## Attaching package: 'rstatix'

## The following object is masked _by_ '.GlobalEnv':
##
##      is_outlier

## The following object is masked from 'package:stats':
##
##      filter

##
# Get help
##

?tukey_hsd()

```

Perform Tukey HSD analysis and display the results

```

tukey_hsd(model_spray)

## # A tibble: 15 x 9
##   term  group1 group2 null.value estimate conf.low conf.high  p.adj
##   * <chr> <chr> <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 spray A      B          0    0.858    -3.91     5.63 9.95e- 1
## 2 spray A      C          0   -12.5    -17.3    -7.77 1.28e- 9
## 3 spray A      D          0    -9.72   -14.5    -4.95 1.47e- 6
## 4 spray A      E          0   -11.0   -15.8    -6.27 5.64e- 8
## 5 spray A      F          0     2.56    -2.21     7.33 6.19e- 1
## 6 spray B      C          0   -13.4   -18.2    -8.63 1.37e-10
## 7 spray B      D          0   -10.6   -15.4    -5.81 1.77e- 7
## 8 spray B      E          0   -11.9   -16.7    -7.13 6.5 e- 9
## 9 spray B      F          0     1.7    -3.07     6.47 9.01e- 1
## 10 spray C     D          0     2.82    -1.96     7.59 5.16e- 1
## 11 spray C     E          0     1.50    -3.27     6.27 9.39e- 1
## 12 spray C     F          0    15.1     10.3    19.9 0
## 13 spray D     E          0    -1.32    -6.09     3.46 9.65e- 1
## 14 spray D     F          0    12.3     7.51    17.1 2.46e- 9
## 15 spray E     F          0    13.6     8.83    18.4 7.94e-11
## # i 1 more variable: p.adj.signif <chr>

```

Notice what does `tukey_hsd()` produces and automatically displays. It shows the means that are subtracted (Group 2 - Group 1), the difference in those means (labeled “estimate”), the lower and upper ends of the 95% CI (the default) and the adjusted p-value for Tukey’s HSD. The order, top to bottom, in the table is maintained throughout everything that we will see below. That make interpreting the graphical display that

we will ultimately produce easy. Note that ns = not statistically significant.

From an examination of this table, what conclusion do you draw about the relative efficacy of the treatment pairs B and A, F and A, F and B, D and C, E and C, and E and D?

Please place your answer in the space below.

**B and A are not significantly different**

**F and A are not significantly different**

**F and B are not significantly different**

**D and C are not significantly different**

**E and C are not significantly different**

**D and E are not significantly different**

**none of these pairs have significantly different results between them.**

For simplicity in creating a visual display of these results, we will use the core R functions and the installed libraries to redo this analysis. First, we will rerun our ANOVA using “aov()” and display the results. Then we will redo the post hoc analysis using Tukey’s HSD.

```
Test <- aov(lm(count ~ spray, data = Insect_Spray))
```

Test

```
## Call:
##   aov(formula = lm(count ~ spray, data = Insect_Spray))
##
## Terms:
##              spray Residuals
## Sum of Squares 2785.807 1046.826
## Deg. of Freedom      5      66
##
## Residual standard error: 3.982587
## Estimated effects may be unbalanced
```

Now use the function that performs Tukey’s HSD, then display the results. Please check some of the results under “diff” to be sure you know how they are calculated. You will need to use the table of means that we created above.

Remember the order of the results in this table will be the same as the order of the graphical results we will produce below

```
posthoc_analysis <- TukeyHSD(Test, 'spray')
```



Display the results

```
posthoc_analysis
```

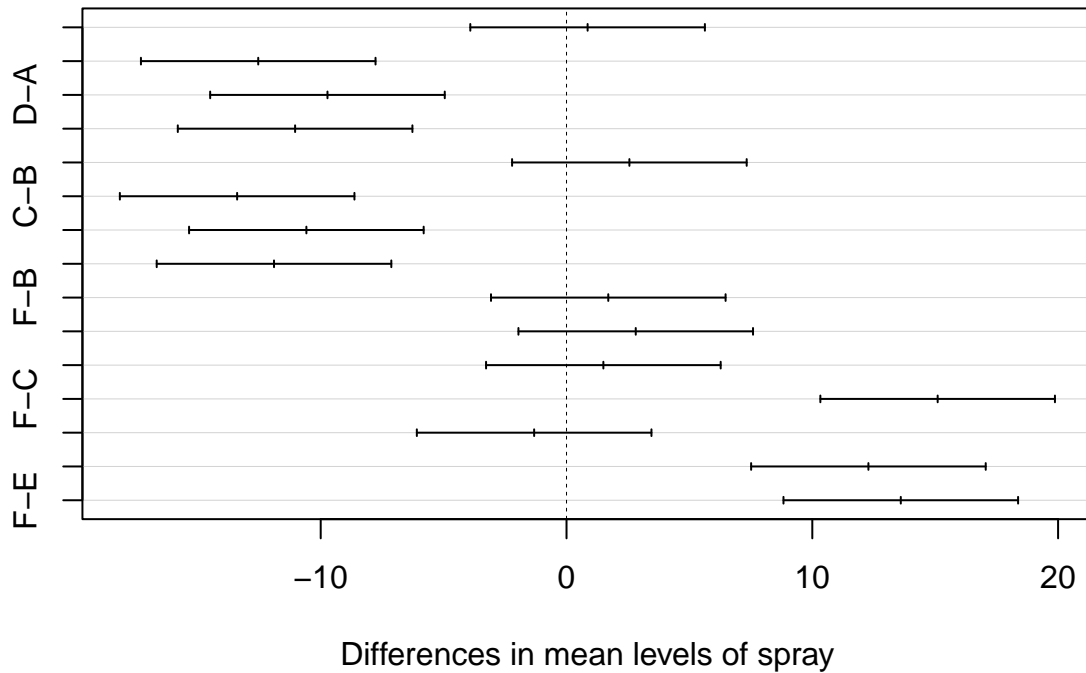
```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = lm(count ~ spray, data = Insect_Spray))
##
## $spray
##      diff      lwr      upr    p adj
## B-A  0.8583333 -3.913791  5.630458 0.9948498
## C-A -12.5416667 -17.313791 -7.769542 0.0000000
## D-A  -9.7250000 -14.497124 -4.952876 0.0000015
## E-A -11.0416667 -15.813791 -6.269542 0.0000001
## F-A   2.5583333 -2.213791  7.330458 0.6185879
## C-B -13.4000000 -18.172124 -8.627876 0.0000000
## D-B -10.5833333 -15.355458 -5.811209 0.0000002
## E-B -11.9000000 -16.672124 -7.127876 0.0000000
## F-B   1.7000000 -3.072124  6.472124 0.9005761
## D-C   2.8166667 -1.955458  7.588791 0.5158564
## E-C   1.5000000 -3.272124  6.272124 0.9393173
## F-C  15.1000000 10.327876 19.872124 0.0000000
## E-D  -1.3166667 -6.088791  3.455458 0.9647441
## F-D  12.2833333  7.511209 17.055458 0.0000000
## F-E  13.6000000  8.827876 18.372124 0.0000000
```

#### Visualization

The following “plot()” function produces a plot of the results from the function TukeyHSD()” using “aov()”.

```
plot(posthoc_analysis)
```

### 95% family-wise confidence level



The graph displays the 95% CI (the default value) about the mean difference in growth rates for each treatment group pair; in the same order in which the differences in pairs appeared in the tables above.

Those CI's that are different from zero (i.e., those CI's that do not cross the dashed vertical line at 0.0) are the ones of interest.

The vertical line at zero indicates no difference in the means. Notice that these CI's are for the same pairs as the test concluded that there was no statistically significant difference in treatment effect.

The graph clearly shows those pairs that are the most different. notice the result for the pair F-C, which is the horizontal line that is fourth from the bottom. Earlier we deduced that F was the control treatment (no insecticide spray).

Please review the boxplots we did while examining this visualization of the results of Tukey's HSD. You should be able to see that they are both showing the same conclusions. That should be comforting!

Hey, maybe this statistics stuff is not all black magic!!

Looking at this result as displayed in the graph, and from the data displayed in the tables above, what do you conclude about the efficacy of C versus that of all of the other insecticides?

Please place your answer in the space below.

Given that spray C has the lowest mean count, and the highest difference

of mean abundance between sprays, we can say that C is the most effective

The End of Assignment #7.