Stat450 Biosolid Short Report

Harry Xu

2018-02-7

Summary

This project report aims to analyse the data collected from an experiment conducted by soil scientists to study the effect of biosolids. The main interest is to find how and whether the biosolid treatment has an effect on the plant health. The main statistical analysis method used is the Analysis of Variance. The results show that the average plant coverage for different species are significanlty different for the control group, and there is a significant treatment effect on different species.

Introduction

This report forms a statistical analysis based on a complete randomized experiments conducted in Jesmond, BC by Soil Scientists at UBC, aiming to examine the long term impacts of biosolids application on grassland plant communities and soil health from 2002 to 2016. The average of individual plant species coverage value in each block for each treatments were recorded and are the main response vaiables for this analysis. Specificaly, the main objectives of this analysis are, to first dicusses whether plant coverage value differ among spcies in the control group, and second, whethere the biosolid treatment is the same for all species. To carry out these obejectives, the Analysis of Variances methods(ANOVA) are used to test these hypothesis and demonstrate the significance of the differences, if any. This short report will contain two main sections, Results and Conclusions, here is the layout for both sections:

Results: -Data Description -Statistical Methods and Assumptions -Data Analysis -Figures Conclusions: -Summary Results and Interpretaion -Future Recommendations

Results

Data Description

The data(32 by 4) has n=32 observations, each observation is categorized by Species (ASAG, HECO, KOMA or LITT), Block (1,2,3 or 4), Treatment (Biosolids or Control), and each observation is an average corresponding plant coverage value.

Statistical Method and Assumptions

The main statistical methods discussed in this report are One Way and Two Way ANOVA test. The ANOVA test method is a hypothesis test used to study the effect of one or more qualitative variables (factors) on a quantitative outcome variable. For all ANOVA test, the main assumption is that the population variance for all groups are the same.

Specifically, One Way ANOVA was used to find whether all species(without biosolids treatment) are equally effective. Here, the varaible we are interested in is plant species. Two Way ANOVA was used to test the effect of the biosolid in different species because we are interested in more than just one explanatory variable. There are two varaibles of interest: Species and Treatment. In a Two Way ANOVA, interaction effect may appear to be significant depending on the data. An interaction effect occurs when the impact of one variable(factor) depends on the level(s) of another.

On a side note, a two-sample T-test's result is equivalent to One Way Anova when there are two groups. Therefore, sample T-tests is helpful to find whether the two population mean for a specific group's value is

significantly differnt from the other. In R, we used TukeyHSD as it has an advantage by incorporating an adjustment for sample size that produces sensible intervals for mildly unbalanced designs.

Data Analysis

The two hypothesis we want to test in mathemetical forms are: 1. Do all species grow equally in the area without treatment? $H_0 : \mu_{ASAG} = \mu_{HECO} = \mu_{KOMA} = \mu_{LITT}$ $H_a : At\ least\ one\ of\ the\ species\ population\ mean\ is\ different.$

2. Is the effect for biosolid treatment the same for all species? $H_0 : \beta_{ASAG} = \beta_{HECO} = \beta_{KOMA} = \beta_{LITT}$ $H_a : At\ least\ one\ of\ the\ species\ treatment\ effect\ for\ plant\ coverage\ values\ is\ different.$

For the both hypothsis, we assume that all group population variances are the same. We use One Way ANOVA function aov() built in R to test the first hypothesis on whether the 4 groups have the same population mean. The test statistic for ANOVA is F, and it compares the between group variation with the within group variation ratio $F = \frac{MSB}{MSW}$. A large F value here(25.32) has a small significant p-value (1.78e-05) indicates that the between group variation is much larger than the within group variation, contrary to our hypothsis that all group means are equal. Therefore, there is strong evidence to reject the null hypothesis at a significance level of 5%. In other words, there is a strong evidence that not all species in the control gorup have the same mean plant coverage value, or, at least one of the species has a different group mean. The control group's data can be visualized in a box plot in Figure 1. Below is the summary table from the ANOVA test from the R console:

```
biosolid <- read.csv("soil_blocks.csv", header = T)
controldata <- subset(biosolid, biosolid$Treatment == "Control")
(summary(anova <- aov(y.avg ~ Species, data = controldata)))
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Species       3 2964.0     988   25.32 1.78e-05 ***
## Residuals    12  468.3      39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Although we found strong evidence that at least one of the spcies' population mean coverage is different from others, we might also want to find which two groups' means are significantly different from each other to obtain more information. We used a paired-sample t-test, as the result is equivalent to a One Way Anova when there are two groups. The function for this is TukeyHSD( ).

```
(TukeyHSD(anova))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = y.avg ~ Species, data = controldata)
##
## $Species
##                diff       lwr       upr    p adj
## HECO-ASAG   26.5875  13.472838  39.70216 0.0003047
## KOMA-ASAG    2.6125 -10.502162  15.72716 0.9327584
## LITT-ASAG   30.1125  16.997838  43.22716 0.0000950
## KOMA-HECO  -23.9750 -37.089662 -10.86034 0.0007621
## LITT-HECO    3.5250  -9.589662  16.63966 0.8540285
## LITT-KOMA   27.5000  14.385338  40.61466 0.0002236
```

From the above table, we found the paired species coverage means that are significantly different at a 5% significance level based on the p-values are: HECO-ASAG, LITT-ASAG, KOMA-HECO, and LITT-KOMA.

For the second hypothesis, We use Two Way ANOVA function aov() built in R to test whether the biosolid treatment effect is the same for all species. Again, we use ANOVA function aov() built in R to carry out this

test as shown in table.

```
(summary(aov(y.avg ~ Species*Treatment, data = biosolid)))
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## Species           3  18187    6062  160.59 5.14e-16 ***
## Treatment         1   1258    1258   33.33 5.97e-06 ***
## Species:Treatment 3   4716    1572   41.64 1.14e-09 ***
## Residuals        24    906      38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the table above, we see that by looking at the p-values, the main effect for Species, for Treatment, as well as the Species*Treatment interaction effect are all significant at the 5% significant level. The main effect of a variable is the effect of this variable have on the response variable, while ignoring the effect of all other variables. In this context, Species has an siginificant main effect, meaning that each species is significant different in plant coverage value. Treatment also has a significant main effect, meaning there is a significant difference between in plant coverage value between the Control Group and the Biosolid Treatment group. Finally, the significant interaction effect between Species and Biosolid Treatment suggests that among all species, the biosolid treatment effect might be significantly different from one species to another species, or, for the two different treatments, each species has a siginifiant different mean for average plant cover value. The visualization of the siginificance are shown in Figure 2 and Figure 3 of the Figures Section.

Figures

Figure 1: Box plot for the Control Group

```
boxplot(y.avg ~ Species, data = controldata, ylab = "Avg Plant Coverage")
```
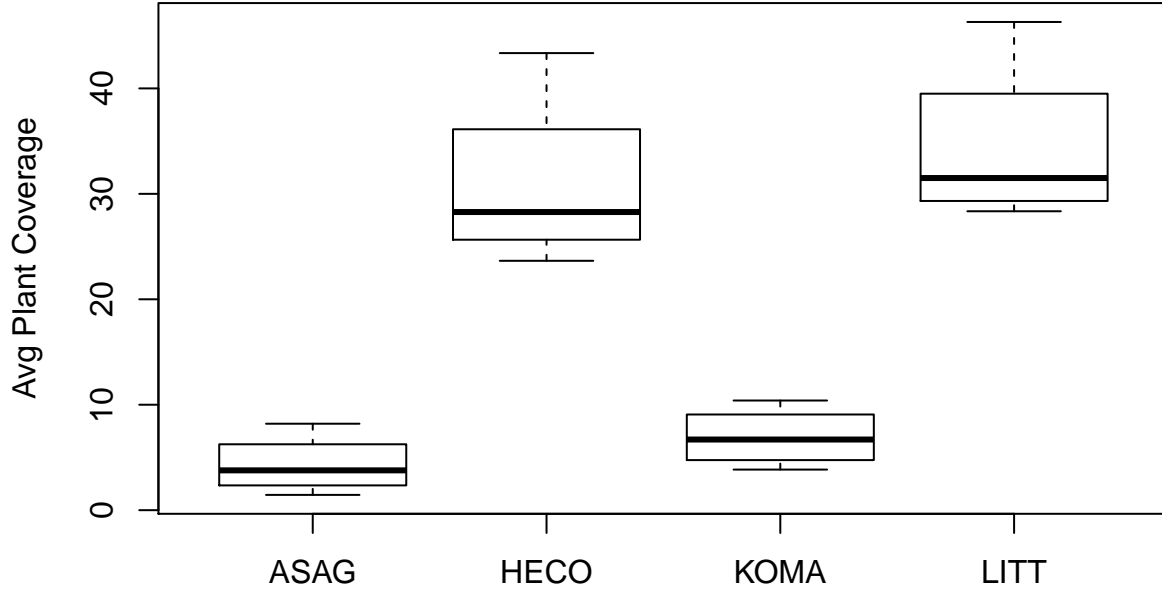

Figure 2: Interaction plot among groups

```
interaction.plot(biosolid$Species, biosolid$Treatment, response = biosolid$y.avg, ylab = "Avg Plant Cove
```
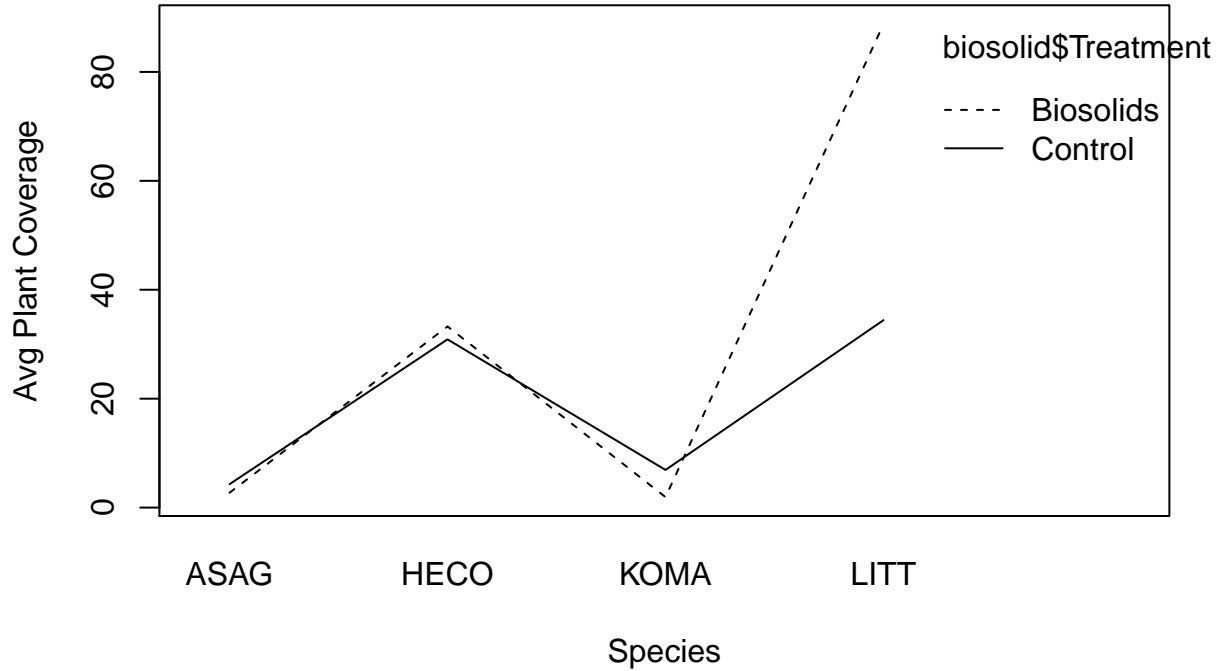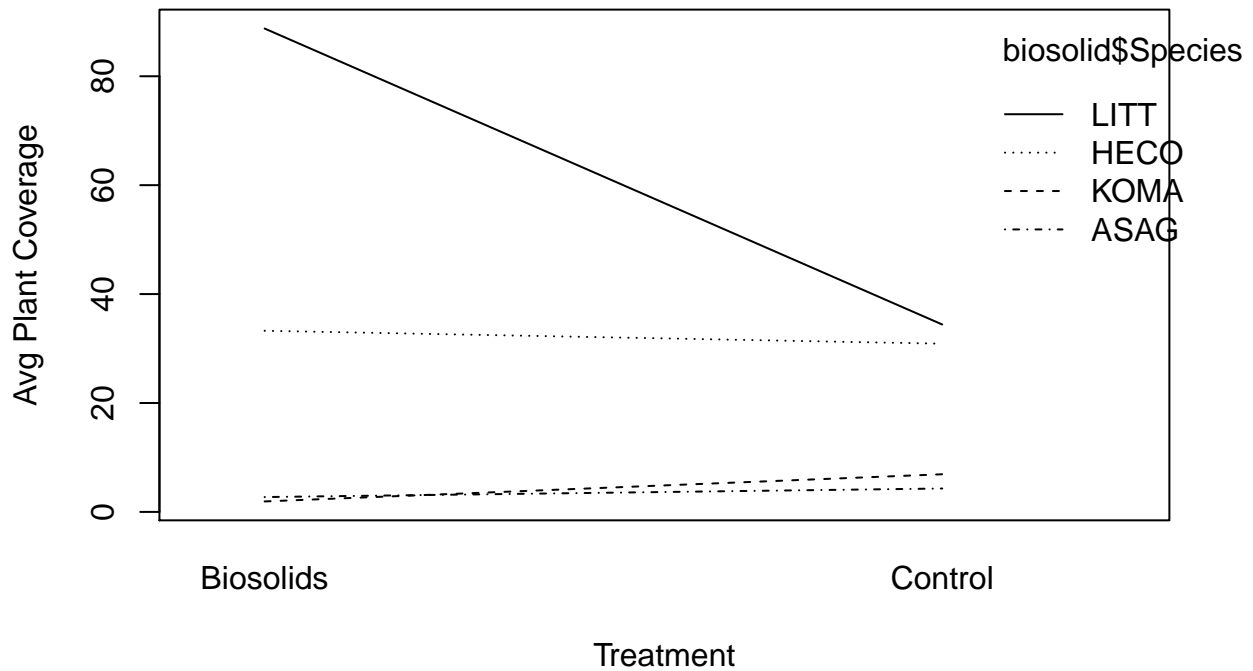
Figure 3: Interaction plot between treatments

```
interaction.plot(biosolid$Treatment, biosolid$Species, response = biosolid$y.avg, ylab = "Avg Plant Cove
```

Conclusions and Recommendations

Summary Results and Interpretaion

In summary, we are able to address the two hypothesis at a 5% significance level. We discovered that the

four different species in the Control group, ASAG, HECO, KOMA, and LITT have different average plant coverage value. In particular, HECO-ASAG, LITT-ASAG, KOMA-HECO, and LITT-KOMA are significantly different. Then, we find that the biosolid treatment affect the species differetly.

Future Recommendations

Although plant coverage value is a great indicator for soil health in this experiment, it is also possible that other confunding factors may also affect the difference in the species group means, as the expereiment was conducted outdoors in a farm, and has a long time span for over 12 years. The result of the analysis can be imporoved by conducting such expereiment in an isolate environment such as inside a lab, where levels of other confonding factors can be controled.