STAT 450 - Biosolids Application Case Study

David Yin

Feb 5, 2018

Summary

The client is interested in examining long term impacts of biosolids application on grassland plant communities and soil health for his masters thesis in soil science. Soil plant coverage data were taken from a field experiment in northern BC in various different periods between 2002 and 2016. A one-way ANOVA test was used to determine whether mean plant coverage were the same across all four species tested, before the application of any biosolids. The results show that the species did not all have the same mean plant coverages before any biosolids application. In particular, species HECO and species LITT's plant coverage values were different compared to species ASAG and species KOMA's. In addition, a two-sample t-test was used to determine which species, if any, had different mean plant coverage before and after the application of biosolids to the soil. The results show that the application of biosolids resulted in a change in mean plant coverage for species KOMA and LITT, but did not result in a change for species ASAG and HECO.

Introduction

The use of biosolids in soil is theorized to improve soil structure by providing nutrition for microbial activity, which lead to the formation of soil aggregates. The improved soil structure can then increase the water and nutrient capacity of the soil and can lead to an incrwease in plant species composition. An experiment can be designed to test this scenario by assessing the plant coverage in these soil before and after biosolids application.

To accomplish this, a field experiment was carried out in 2002 at Jesmond, BC with a randomized complete block design. The four species studied are: ASAG, HECO, KOMA, and LITT. Four blocks of soil were studied for each species. Two treatments were replicated in each of the four blocks, with the first one having no biosolids application (the control) and the second one having biosolids application at 20 dry Mg ha-1. Each block was 50m by 50m in size and protected with fencing. Sample plant cover values were collected by a visual assessment of each of the blocks of soil. Samples were collected before biosolids application in 2002, and for four consecutive years following application, and again in 2016.

Based on the questions the client was interested in investigating in, we came up with the following two hypotheses to address the client's needs in his study:

1. **Whether the mean plant coverage is the same for all four species prior to the application of biosolids**, and
2. **Whether there was a difference in mean plant coverage before and after biosolids application for the different species**.

In the following results section, we delve deeper into the details of the data available, the analysis tools used, and the results of these analyses.

Results

Description of Data

The data consisted of 32 rows and 4 columns. Each of the rows was composed of of the species that was sampled, the block number, the treatment (biosolids vs no biosolids), and the block's sample average plant cover value. The following is a snippet of the start of the dataset:

```r
soil = read.csv("soil_blocks.csv")
head(soil)
```

```
##   Species Block Treatment y.avg
## 1    ASAG     1 Biosolids  5.65
## 2    ASAG     1   Control  3.25
## 3    ASAG     2 Biosolids  0.80
## 4    ASAG     2   Control  1.45
## 5    ASAG     3 Biosolids  2.25
## 6    ASAG     3   Control  8.20
```

Statistical Methods Used: Hypothesis 1

There are two hypotheses we want to test. The first hypothesis we want to test is whether or not the mean plant coverage differs across different species before any biosolids were applied. In other words, we have:

$H_0 : \mu_{ASAGcontrol} = \mu_{HECOcontrol} = \mu_{KOMAcontrol} = \mu_{LITTcontrol}$

$H_A : At\, least\, one\, pair\, of\, the\, species'\, population\, mean\, for\, plant\, coverage\, is\, different.$

Where μ is the population mean of plant cover value.

We did this using a one-way ANOVA on the sample average plant coverages of the control species (those that did not get biosolids applied). ANOVA studies the effect of qualitative variables on a quantitative outcome. In our case, the quantitative outcome would be **mean plant cover value**, the factor would be our 4 **species**. In order to proceed with the one-way ANOVA, we assumed that all the group population variances are equal, that they are independently and identically distributed within each group, and that their residuals are normally distributed within each group.

The p-value of a test is the probability of finding the observed results when the null hypothesis (H_0) is true. We can determine if the test rejects the null hypothesis if the p-value is less than 0.05 (using a significance level of 0.05). In the case of ANOVA, the area to the right of the F-statistic, which is a ratio of variation between sample means and variation within samples, in the F-distribution is the p-value.

Later on, we also used two-sample t-test to compare the individual pairs of mean plant cover values. A two-sample t-test takes into account the sample mean and sample standard deviation of the two groups to determine if two population means are equal. This can help us see which specific species have different mean cover value than which other species. The assumptions we made are that the values within each group are randomly sampled, that a reasonably large sample size is used, that the distributions of the residuals are normal, and that each observation is independent from other observations in the same group.

Analysis and Findings: Hypothesis 1

```
soil_control = subset(soil, Treatment=="Control")
mod = aov(y.avg ~ Species, data = soil_control)
summary(mod)
```

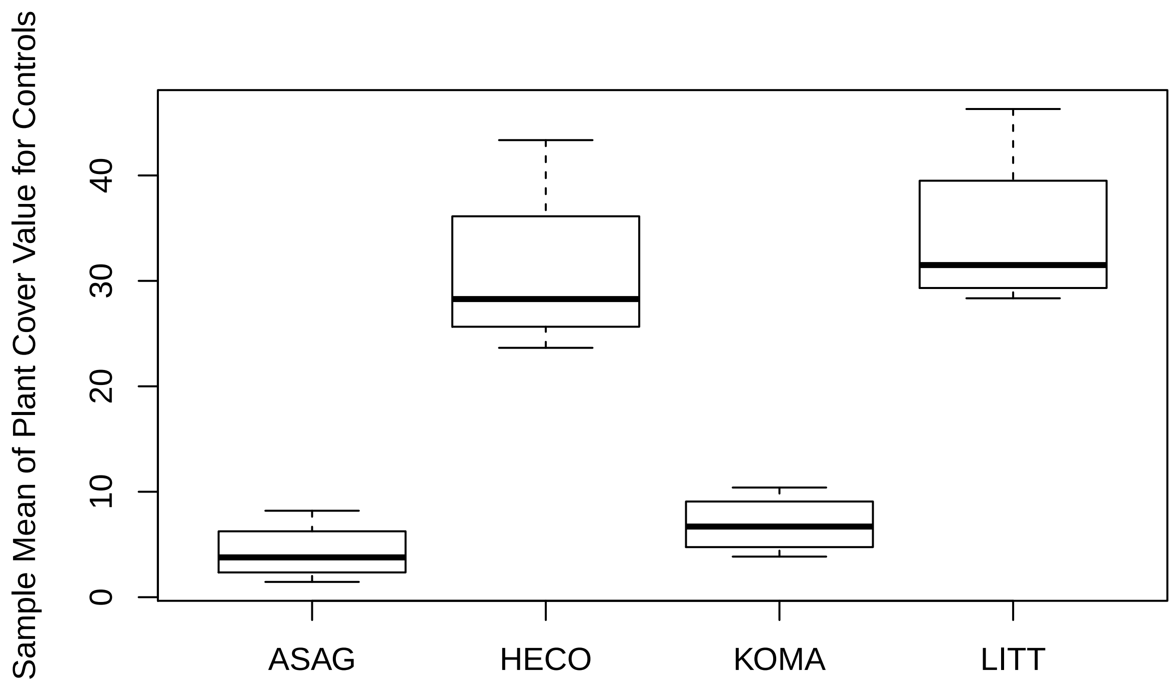
```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Species      3 2964.0     988   25.32 1.78e-05 ***
## Residuals   12  468.3      39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The between group variation can be compared with the within group variation using the F-statistic: $F = \frac{MSB}{MSW}; F \sim \mathcal{F}(K-1, N-K)$. The p-value from this test is the shaded area to the right of the F-statistic under the F distribution.

The F statistic is 25.32. The p-value is $1.78 * 10^{-5}$. This value is smaller than our significance level of 0.05. Thus, since the test's p-value is less than 0.05, we can conclude that the plant coverages were not the same across the different species before the biosolids application at the 5% significance level.

We answered our first hypothesis, but the above result does not tell us much about the underlying plant coverages of each of the individual species. We would like to investigate which one (or more) of pairs of ASAG, HECO, KOMA, and LITT had a significant difference in their plant coverage value. To do this, we started off by creating a boxplot to visualize the coverage values:

```
boxplot(y.avg ~ Species, data = soil_control,
        ylab = "Sample Mean of Plant Cover Value for Controls")
```



From this plot, it appeared that HECO and LITT have higher plant coverage values than ASAG and KOMA. We also looked at the differences of each pair using multiple two-sample t-tests provided by TukeyHSD in R:

```r
TukeyHSD(mod)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = y.avg ~ Species, data = soil_control)
##
## $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 t-test for each pair, we noticed that the pairs that have adjusted p-values less than 0.05 are: HECO-ASAG, LITT-ASAG, HECO-KOMA, and LITT-KOMA. So using a significance level of 0.05, we can conclude that HECO's plant coverage value is significantly different from ASAG and KOMA's, and LITT's coverage value is also significantly different from ASAG and KOMA's.

Statistical Methods Used: Hypothesis 2

The second hypothesis we want to test is whether or not the biosolids treatment had a significant effect on any of the species.

Our hypothesis is, for a given species, the biosolids treatment had no effect on mean plant cover value. In other words, for a given species, the mean plant cover value is the same with or without biosolids treatment:

$H_0 : \mu_{biosolids} = \mu_{control}$

$H_A : \mu_{biosolids} \neq \mu_{control}$

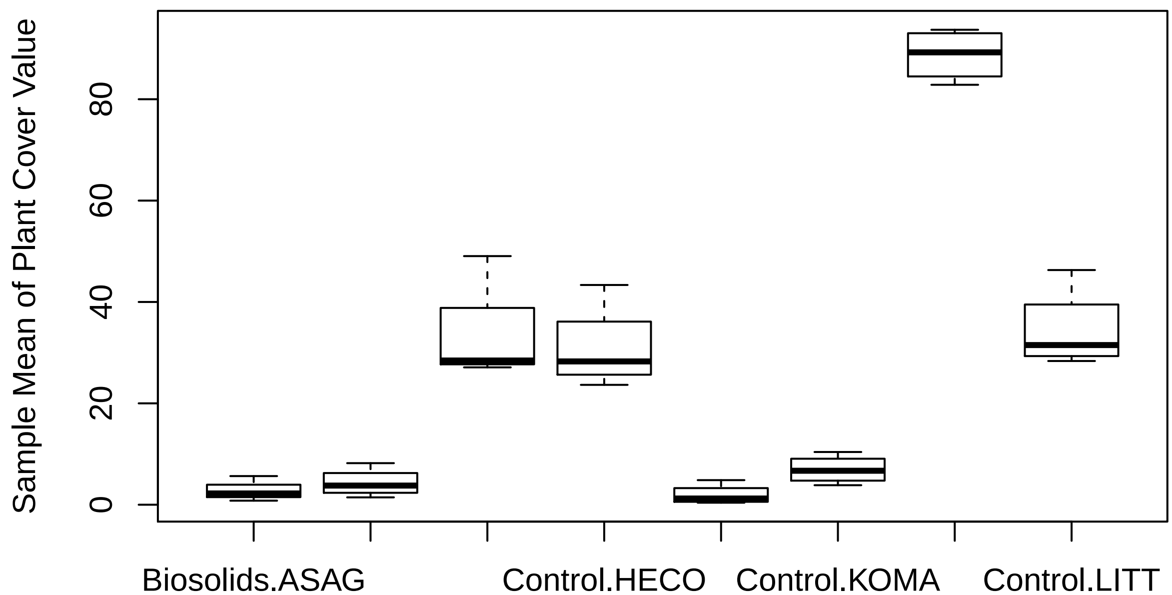
for a given species.

Again, we can do this using a two-sample t-test on groups of biosolids vs control results for each species. This two-sample t-test allows us to examine if a population mean is significantly different from another population mean. We performed this test using the samples we collected from the biosolids treatment and control treatment for each species. Once more, the assumptions we made are that the values within each group are randomly sampled, that a reasonably large sample size is used, that the distributions of the residuals are normal, and that each observation is independent from other observations in the same group.

Analysis and Findings: Hypothesis 2

First, we looked at a boxplot showing the differences of each species before and after control:

```
boxplot(y.avg ~ Treatment:Species, data = soil,
        ylab = "Sample Mean of Plant Cover Value")
```



From this plot, it appeared that the biosolids application had a large effect on LITT's mean plant cover value. We can test if this is the case for every species. The following are the two-sample t-tests on the individual species' control vs biosolids groups to see if the biosolids had an effect on mean plant cover value for each of the species.

```
t.test(y.avg ~ Treatment, data = subset(soil ,Species == "ASAG"))
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  y.avg by Treatment
## t = -0.89457, df = 5.4621, p-value = 0.4087
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.988039  2.838039
## sample estimates:
## mean in group Biosolids   mean in group Control
##                   2.725                   4.300
```

```r
t.test(y.avg ~ Treatment, data = subset(soil ,Species == "HECO"))
```

```
##
##  Welch Two Sample t-test
##
## data:  y.avg by Treatment
## t = 0.349, df = 5.768, p-value = 0.7395
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -14.44023  19.19023
## sample estimates:
## mean in group Biosolids   mean in group Control
##                 33.2625                 30.8875
```

```r
t.test(y.avg ~ Treatment, data = subset(soil ,Species == "KOMA"))
```

```
##
##  Welch Two Sample t-test
##
## data:  y.avg by Treatment
## t = -2.8736, df = 5.4465, p-value = 0.03158
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.3413148 -0.6336852
## sample estimates:
## mean in group Biosolids   mean in group Control
##                  1.9250                  6.9125
```

```r
t.test(y.avg ~ Treatment, data = subset(soil ,Species == "LITT"))
```

```
##
##  Welch Two Sample t-test
##
## data:  y.avg by Treatment
## t = 11.313, df = 5.0664, p-value = 8.686e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  42.04942 66.65058
## sample estimates:
## mean in group Biosolids   mean in group Control
```

```
##                  88.7625              34.4125
```

Here, we see that, for the species ASAG and HECO, the p-values are greater than 0.05, thus for these two species, we do not reject our null hypothesis that the mean plant coverage before and after biosolids application is the same. However, for the species KOMA and LITT, the p-values are less than 0.05. Thus, for these two species at the 5% significance level, we reject our null hypothesis in favor of our alternative hypothesis that the mean plant coverage before and after biosolids application is different.

Conclusions and Possible Future Steps

In conclusion, we addressed two hypotheses the client is interested in: **1. Whether the four mean plant coverage are the same for the controls**, and **2. Whether there was a difference in mean plant coverage before and after biosolids application for the different species**. For the first hypothesis, using one-way ANOVA, we discovered that the mean plant cover value is not the same across all of the species. After doing a two-sample t-test on each pair of species, we also discovered that HECO's plant coverage value and LITT's plant coverage value are significantly different from ASAG and KOMA's. With regards to our second hypothesis, using a two-sample t-test on the mean plant cover values before biosolids application vs after biosolids application, we concluded that the mean plant coverage is not significantly different with or without biosolids for species ASAG and HECO, but is significantly different with biosolids compared to without for species KOMA and LITT.

Some of the possible future steps we recommend to the client include an investigation into the quality of the data collected. It would be wise to look into the method used to obtain the biosolid and control treatments' mean plant cover value. Is there a systematic error involved that can be adjusted? Were there any changes to the land over the time the data was collected? Also, it would be valuable to explore novel data collection techniques such as a neural network-based image recognition algorithm to determine plant cover value, instead of using the human eye for assessment. We hope that the answers we provide to the client will help him or her towards the ultimate research goal.