

Biostatistics Final Project: Does the species and size of the tree matter in the location of SLF eggs?

Abhimaan Mayadam

Background

Spotted Lanternflies are an invasive species that first appeared in Pennsylvania in 2014. Since then, these Lanternflies have been slowly spreading across the United States. They first appeared in Berks County, Pennsylvania, and affected a wide variety of plants, such as hops, hardwood trees and even grapes (PennState Extension). Researchers from Temple University have conducted a multivariable assessment of Spotted Lanternflies and their egg-laying patterns. This assessment documented the location of egg cases within 66 trees within 15 sites, which was narrowed down from the initial 141,984 surveyed sites, 7,363 which had a Spotted Lanternfly population (Ramirez et al 2023). While conducting field surveys at the 15 sites, the researchers noted that there were egg cases high up on some of the trees, making those egg cases inaccessible to them. They did not go into much detail about these inaccessible egg cases. I hypothesize that there is a correlation between these inaccessible egg cases and the DBH or species of the tree.

Methods

The original dataset was collected over the course of two Fall-Spring seasons in the 15 sites, and the 66 trees were monitored from the Fall of 2018 to the Spring of 2020. The field data collected included: Tree Diameter at Breast Height in centimeters, count of the egg cases, the number of eggs in the egg cases, if the egg cases were inaccessible, the time the tree was banded to trap Spotted Lanternflies, if they had to reapply the band, genus of the tree, and the species of the tree. Each collected egg mass, site, and tree had their own unique identifier. The invasion duration of each site was documented from previous surveys and the mean tree cover was calculated via Google Earth. The initial statistical analyses were done in R, like this current statistical analysis. The analyses completed were Spearman's Rank Correlation and a logistic regression for the correlation between the high (or inaccessible) egg cases and the tree diameter at breast height. A random effects ANOVA was conducted between Tree DBH in centimeters and Tree Species to calculate the variation between species. A Fisher's exact test was conducted to calculate the independence of the categorical variables. Nonparametric tests were performed due to data not meeting assumptions, regardless of normalization attempts.

Results

Spotted Lanternflies do not lay their egg cases higher on certain species of tree relative to other species of trees. (Fisher's exact test, $df = 13$, $p < .001$). There is a correlation between inaccessible egg cases laid on trees with a larger DBH. (Spearman's Ranked Correlation, $r_s = -.247$, $n = 1544$, $P < .001$). For every increase in cm for DBH, there is a 4% increase in the prevalence of inaccessible egg cases. (Logistic Regression, $X^2 = 119.38$, $df = 1$, $p < .001$). 74% of the variation in DBH was due to the difference in species. (Random Effects ANOVA, $X^2 = 419.52$, $df = 1$, $p < .001$).

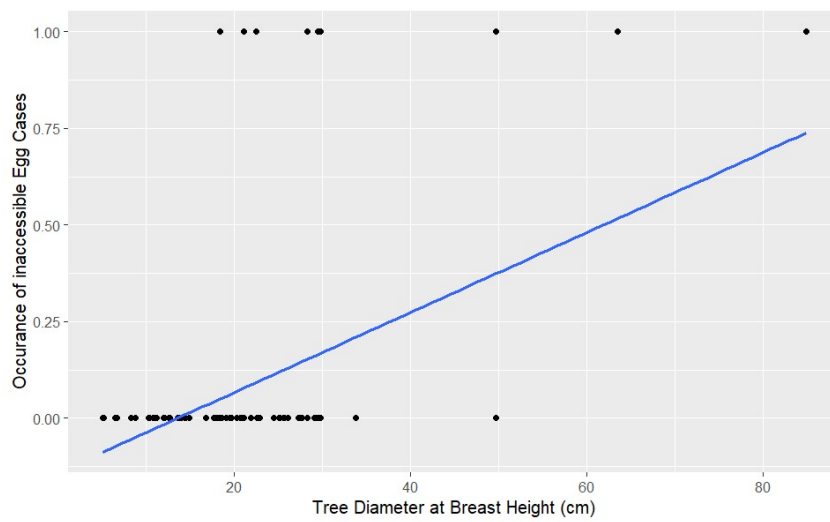


Figure 1 Correlation between Inaccessible eggs and DBH

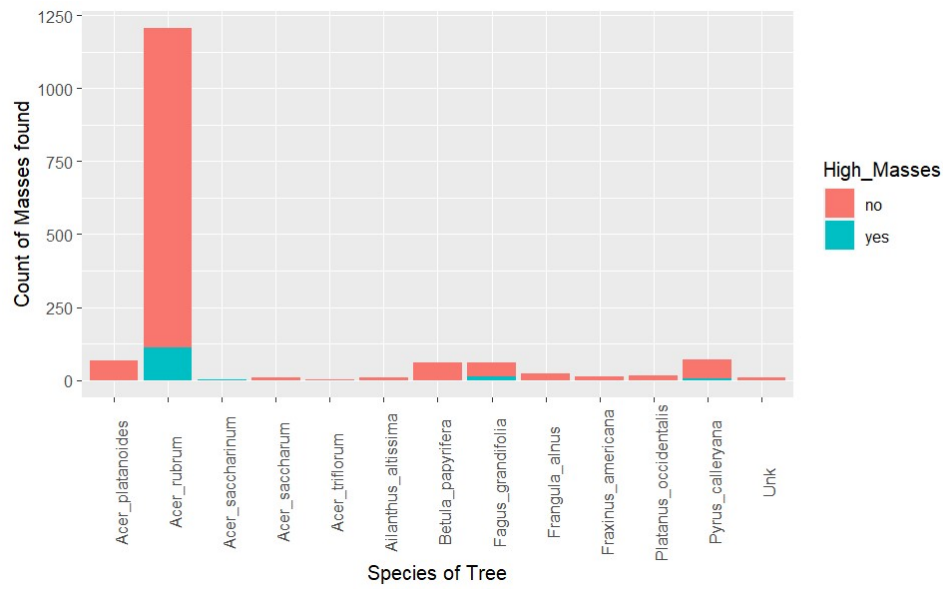


Figure 1 Masses found based on Species of Tree

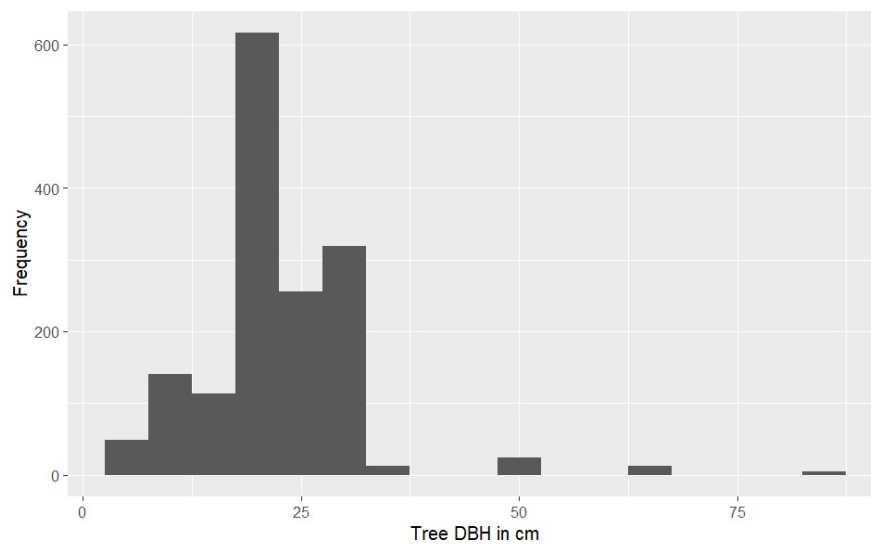


Figure 3 Distribution of Tree DBH

Discussion

This study noted that the Spotted Lanternflies preferred areas with lower canopy cover, such as standalone trees compared to more densely populated trees. This preference in lower canopy cover could be tied to their behavior in determining suitable trees to land on. In Baker et al 2021, the researchers documented the flight patterns of Spotted Lanternflies while flying off of a telephone pole in a vineyard. The researchers postulated that the high contrast between the light-colored sky and the pole confused the Spotted Lanternflies into thinking that it was a tree. The preference for low canopy cover areas could be related to this behavior. As for the data analyzed, Spotted Lanternflies preferred trees with a larger DBH, or a larger trunk. As for a species preference, there was a higher concentration of egg cases found on *Acer rubrum*, or Red Maple trees compared to others such as *Ailanthus altissima*, or the Tree of Heaven. But this higher concentration could be due to the higher abundance of Red Maple in the area. The host tree of the Spotted Lanternfly, *Ailanthus altissima*, was also in this dataset but there were only three trees, and they did not show a preference to them, compared to other trees that had low incidences in the dataset. The slight correlation between DBH and inaccessible tree cases is more likely due to the previously mentioned overabundance of the Red Maple tree within the dataset. Future directions that this study could take is a more controlled study with more documented tree dimensions, such as DBH, tree crown and volume of the trunk in already established populations of Spotted Lanternflies.

References

- Baker TC. Et al. 2021. Visual Responses of Flight-Dispersing Spotted Lanternflies, *Lycorma delicatula* toward a Tall Vertical Silhouette in a Vineyard. *Journal of Insect Behavior* 34:49-60
- Ramirez VA. et al. 2023. Multiscale assessment of oviposition habitat associations and implications for management in the spotted lanternfly (*Lycorma delicatula*), an emerging invasive pest. *Journal of Applied Ecology* 60:3. 411–420. DOI: 10.1101/2022.09.08.507131
<https://datadryad.org/stash/dataset/doi:10.5061/dryad.ffbg79d00>
- Spotted Lanternfly. PennState Extension. <https://extension.psu.edu/spotted-lanternfly>

Appendix A

Statistical Exploration

```
slf <- read.csv("fecundity_data_full_v1_3.csv") #load in csv
mean(na.omit(slf$Tree_DBH_cm)) #average of DBH of tree
median(na.omit(slf$Tree_DBH_cm)) #median of DBH of tree
var(na.omit(slf$Tree_DBH_cm)) #varaince of DBH of tree
sd(na.omit(slf$Tree_DBH_cm)) #standard deviation of DBH of tree
range(na.omit(slf$Tree_DBH_cm)) #range of DBH of tree
```

```
IQR(na.omit(slf$Tree_DBH_cm)) #interquartile range of DBH of tree
100 * (sd(na.omit(slf$Tree_DBH_cm)) / mean(na.omit(slf$Tree_DBH_cm)))
#coefficient of variation
```

Appendix B

Statistical Analysis:

```
#READING DATA
slf <-read.csv("fecundity_data_full_v1_3.csv")
slf_naomit <- na.omit(slf)

#Attempts to normalize and transform the data
shapiro.test(slf_naomit$Tree_DBH_cm)
shapiro.test(slf_naomit$Number_covered_masses_seen)
slf_naomit$Number_covered_masses_seen_log <-
log(slf_naomit$Number_covered_masses_seen)
slf_naomit$Tree_DBH_cm_log <- log(slf_naomit$Tree_DBH_cm)
shapiro.test(slf_naomit$Tree_DBH_cm_log)
shapiro.test(slf_naomit$Number_covered_masses_seen_log)

#CORRELATION
cor.test(slf_naomit$Number_covered_masses_seen,
slf_naomit$Tree_DBH_cm, method = "spearman")# Correlation for Height
and DBH

#LOGISTIC REGRESSION
slf_naomit$binomhigh <- ifelse(slf_naomit$high_masses == "y", 1, 0)
#Binomial Preparation
high_dbh.glm <- glm(binomhigh ~ Tree_DBH_cm, data = slf_naomit, family
= binomial(link = logit)) #GLM for Height and DBH Logistic Regression
summary(high_dbh.glm)
anova(high_dbh.glm, test = "Chi") #Logistic Regression

#RANDOM EFFECTS ANOVA
library(nlme)
slf_species_treedbh <- lme(fixed = Tree_DBH_cm_log ~ 1, random =
~1|Tree_Species, data = slf_naomit) #Alt hypothesis for Random Effects
ANOVA
library(car)
leveneTest(data = slf_naomit, Tree_DBH_cm ~ Tree_Species, center =
mean) #homogeneity of Variance
```

```

shapiro.test(resid(slf_species_treedbh)) #Shapiro Test for Normality
slf_species_treedbh_gls <- gls(Tree_DBH_cm_log ~ 1, data = slf_naomit,
method = "REML") #Null hypothesis for Random Effects ANOVA
anova(slf_species_treedbh, slf_species_treedbh_gls) #ANOVA
varcomp <- VarCorr(slf_species_treedbh) #Calculating Variance
Components
varcomp
varamong <- as.numeric(varcomp[1,1])
varwithin <- as.numeric(varcomp[2,1])
repeatability <- varamong / (varamong + varwithin)
repeatability

#FISHER'S EXACT TEST #Contingency Table Made in Excel, then imported

cont <-data.frame(Yes = c(Contingency$`Count of Yes`), No =
c(Contingency$`Count of No`), row.names = c(Contingency$Tree_Species))
#making Contingency Table
fisher.test(cont, workspace = 2e9) #Fisher's Exact Test, workspace =
2e9 was due to a small workspace size error. The maximum limit of the
workspace is tied to RAM, but I do not know how much it needs, but 2e9
worked on my computer, but it did not for someone on Stack Overflow.

library(ggplot2)
ggplot(slf_naomit, aes(x=Tree_DBH_cm, y = binomhigh)) + geom_point() +
geom_smooth(method = glm, se = FALSE, methods = list(family =
binomial)) + labs(x= "Tree Diameter at Breast Height (cm)", y =
"Occurance of inaccessible Egg Cases")#

ggplot(data = slf_naomit, aes(x=Tree_Species, fill = High_Masses))
+geom_bar(stat = "count") + theme(axis.text.x =
element_text(angle=90)) + xlab("Species of Tree") + ylab("Count of
Masses found")

ggplot(data = slf_naomit, aes(x=Tree_DBH_cm)) +
geom_histogram(binwidth = 5) + labs(x="Tree DBH in cm", y =
"Frequency")

```