A4_20019803

Kai Ellis

February 5, 2020

Hypothesis: (H2) Galerucella damage was higher at sites where the biocontrol agents had been deliberately released than sites where they had not. (H3) Damage by beetles reduced the local abundance of L. salicaria

1. A code chunk loading the required packages.

```
## Warning: As of rlang 0.4.0, dplyr must be at least version 0.8.0.
## x dplyr 0.7.8 is too old for rlang 0.4.3.
## i Please update dplyr to the latest version.
## i Updating packages on Windows requires precautions:
##   <https://github.com/jennybc/what-they-forgot/issues/62>
```

```
##
## 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
```

```
## Warning: package 'ggpubr' was built under R version 3.5.3
```

```
## Loading required package: magrittr
```

```
## Warning: package 'ggfortify' was built under R version 3.5.3
```

```
## Warning: package 'car' was built under R version 3.5.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 3.5.3
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

Note: I am using ggpubr for making the composite plot rather than cowplot. Cowplot's installation was giving me alot of problems, and ggpubr's ggarrange function proved intuitive for creating composite graphs.

2. Code importing the dataframe and check it with the single most useful data-checking function.

```
loosDat <- read.csv("/Users/kelli/Desktop/Biol343/Assignment_4/loosestrife.csv")

summary(loosDat)
```

```
##       site         latitude       longitude         region
##  B1.PLD : 1   Min.   :42.58   Min.   :-81.52   Central:20
##  B2.NUM : 1   1st Qu.:44.10   1st Qu.:-79.83   East   :18
##  B3.LDY : 1   Median :44.56   Median :-79.27   Other  :14
##  B4.SCUG: 1   Mean   :44.41   Mean   :-78.40
##  B5.CAL : 1   3rd Qu.:44.73   3rd Qu.:-76.25
##  B6.DUF : 1   Max.   :45.74   Max.   :-75.67
##  (Other):46
##        release     stemsper50m2     stemheight_cm    inflorescences
##  Nonrelease:32   Min.   :   0.0   Min.   : 42.0   Min.   : 0.100
##  Release   :20   1st Qu.:  24.5   1st Qu.:106.8   1st Qu.: 2.450
##                  Median :  85.0   Median :121.0   Median : 4.400
##                  Mean   : 135.8   Mean   :119.4   Mean   : 4.817
##                  3rd Qu.: 171.8   3rd Qu.:138.0   3rd Qu.: 5.896
##                  Max.   :1000.0   Max.   :171.0   Max.   :13.350
##
##      damage
##  Min.   :0.3464
##  1st Qu.:2.0282
##  Median :2.3726
##  Mean   :2.5564
##  3rd Qu.:3.1445
##  Max.   :4.4505
##
```

```
#I ran a density transformation here so it wouldn't be forgotten later in the assignm
ent.
loosDat <- loosDat %>% mutate(logDense = log10(1 + stemsper50m2))
```

```
## Warning: The `printer` argument is deprecated as of rlang 0.3.0.
## This warning is displayed once per session.
```

3. Code that makes a graph to evaluate the hypothesis that damage is higher at release than nonrelease sites (H2).

```
sumStat  <- loosDat %>% group_by(release) %>% summarise(m = mean(damage, na.rm = T),
                                        sd = sd(damage, na.rm = T),
                                        n = sum(!is.na(damage)),
                                        se = sd/sqrt(n),
                                        LCI = m - se*(qt(0.975,df = (n-1) )),
                                        UCI = m + se*(qt(0.975,df = (n-1) )))
ggplot(data = sumStat) +
  geom_point( aes(y = m, x = release), size = 5, colour = "red" ) +
  geom_errorbar(aes(ymax = UCI, ymin = LCI, x = release ), width = 0.1, lwd =1, colou
r = "red") +
  ylab("Damage index") +
  scale_x_discrete (name ="Release status") +
  ylim(0,3.5) +
  theme_minimal()
```
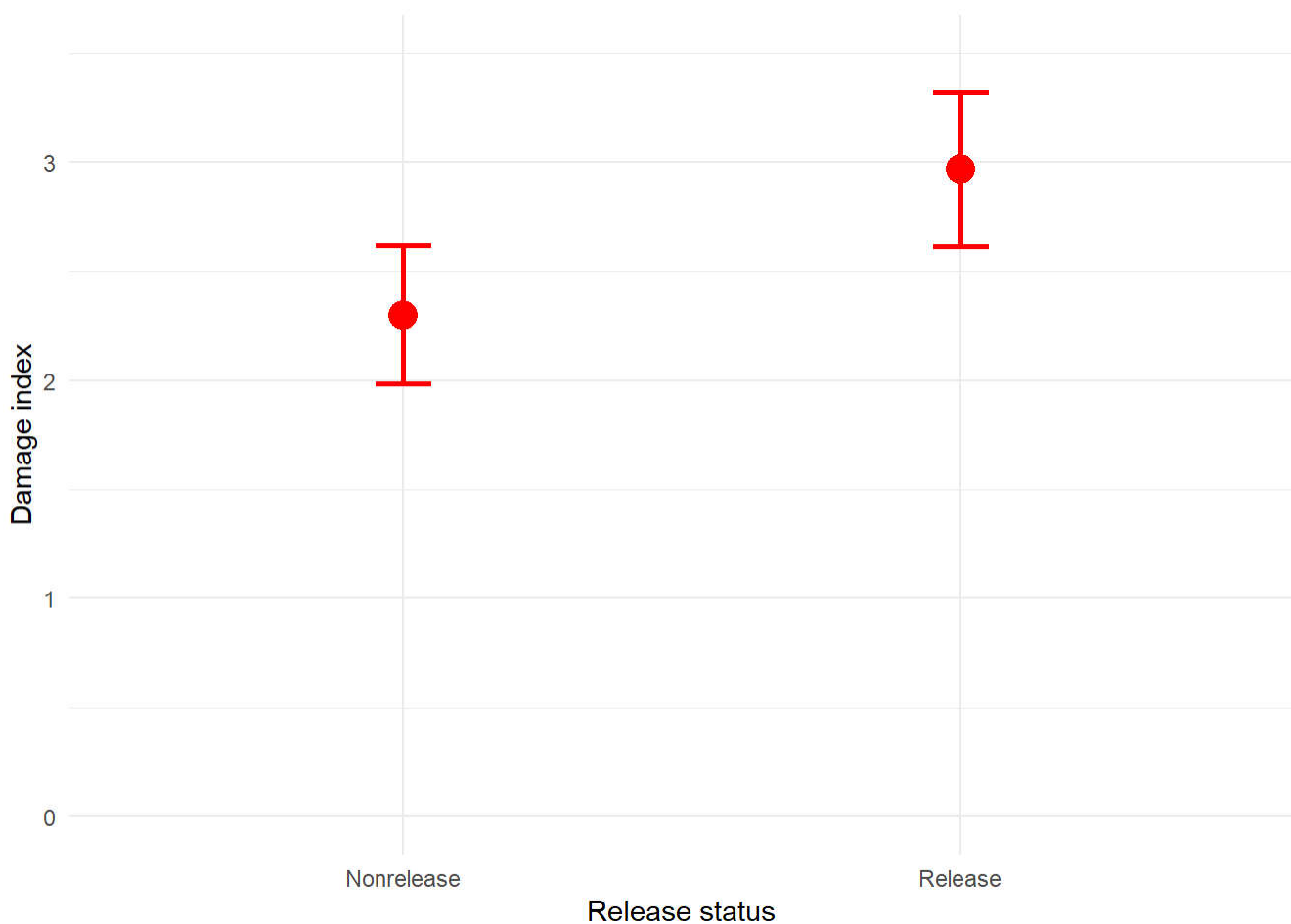

Figure 1. The mean and 95% confidence bars for beetle damage index at sites where Galerucella was and was not purposely released.

4. Code that executes a linear model testing for a difference in mean damage between release and nonrelease sites (H2).

```
lmod <- lm(damage ~ release, loosDat)
summary(lmod)
```

```
##
## Call:
## lm(formula = damage ~ release, data = loosDat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.95296 -0.58033  0.01847  0.57796  2.14211
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.2993     0.1472  15.620  < 2e-16 ***
## releaseRelease    0.6683     0.2374   2.816  0.00695 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8327 on 50 degrees of freedom
## Multiple R-squared:  0.1369, Adjusted R-squared:  0.1196
## F-statistic: 7.928 on 1 and 50 DF,  p-value: 0.006947
```

5. Code and a figure testing the normality of residuals assumption for H2.

```
H2n <- ggplot(data.frame(residuals(lmod))) +
  geom_histogram(aes(x = residuals(lmod), y = ..density..), binwidth = 0.4) +
  labs(x = "Residuals", y ="Density") +
  stat_function(fun= dnorm, args= list(mean= mean(residuals(lmod)),
                                        sd = sd(residuals(lmod))),colour = "blue", lwd
= 2) +
  theme_minimal()

#it is worth noting that at higher values of n(), shapiro-wilks tests (SWT) will dete
ct any minor difference from normality, and can result in false rejections of the nul
l hypothesis
shapiro.test(residuals(lmod))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(lmod)
## W = 0.98975, p-value = 0.9323
```

6. Code and a figure testing the independence of residuals assumption for H2.

```
loos.lmod <- loosDat %>% mutate(resids = residuals(lmod), pred = fitted(lmod))

H2i <- ggplot(loos.lmod, aes(x = release, y = resids)) +
  geom_boxplot() +
  labs(x = "Release status", y ="Residuals") +
  theme_minimal()

leveneTest(resids~release, data = loos.lmod)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1   0.198 0.6582
##       50
```

```
secfigure <- ggarrange(H2n,H2i,ncol = 2, nrow = 1, labels = c("A","B"))
secfigure
```
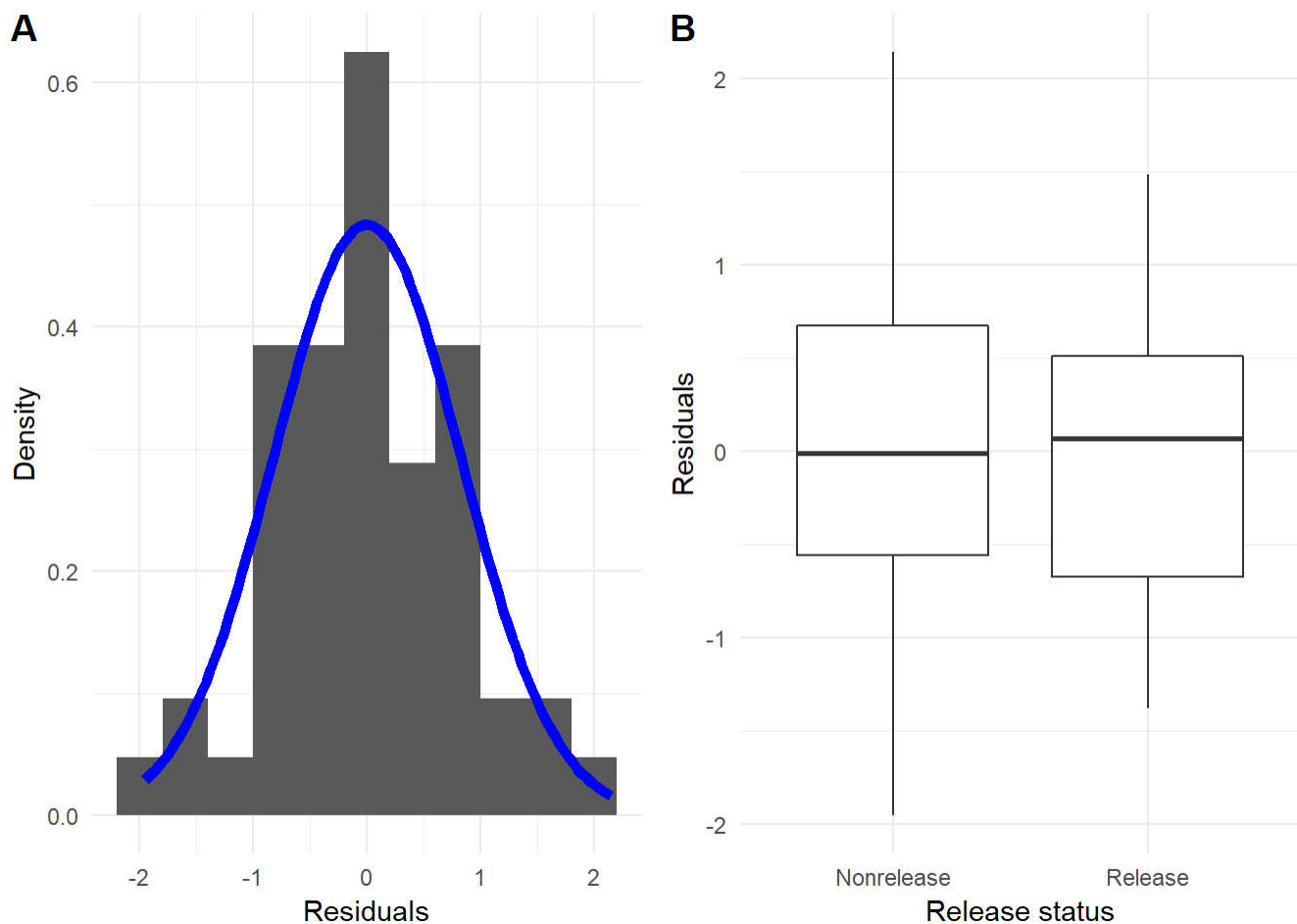

Figure 2. Assessment of the assumptions that residuals of the multivariate data between beetle damage and release site were A) normally distributed, and B) homoscedastic.

7. A permutation test for H2.

```
sum.lmod <- summary(lmod)
f.obs <- as.numeric(sum.lmod$fstatistic[1] )

nreps  <- 1000
frand <- numeric(nreps)
for (i in 1:nreps){
  rdamage <- sample(loosDat$damage, length(loosDat$damage), replace = F)
  sumrmod <- summary(rmod <- lm(rdamage~release, data = loosDat))
  frand[i] <- as.numeric(sumrmod$fstatistic[1])


}
mean(frand >= f.obs)
```

```
## [1] 0.009
```

8. A section of text interpreting your results with respect to H2. *(H2) Galerucella damage was higher at sites where the biocontrol agents had been deliberately released than sites where they had not.*

Sites where Galerucella was intentionally released had a significant increase in damage index relative to sites without intentional release. This can be seen both in the lack of overlap between the 95% confidence intervals (Figure 1.), and in the significant (p = 0.006947) f-statistic from a general linear model (glm) analysis.

Examining the assumptions of normality and independance of residuals required for glm analysis reveals we cannot reject the null hypothesis that the residuals are both normal (p = 0.93) and independant (p = 0.65).

The data supports H2, with glm analysis revealing biocontrol release has a significant positive relation with damage index.

9. Code that makes a graph to evaluate H3.

```
ggplot(loosDat, aes(y = logDense, x = damage)) + geom_point() + labs(x ="Damage inde
x", y = expression(paste("Stems per ",m^2," (",log[10],")" )))+
  theme(axis.title.y= element_text(size = 9.5)) + geom_smooth(method = lm, se = FALS
E, colour = "blue") + theme_minimal()
```
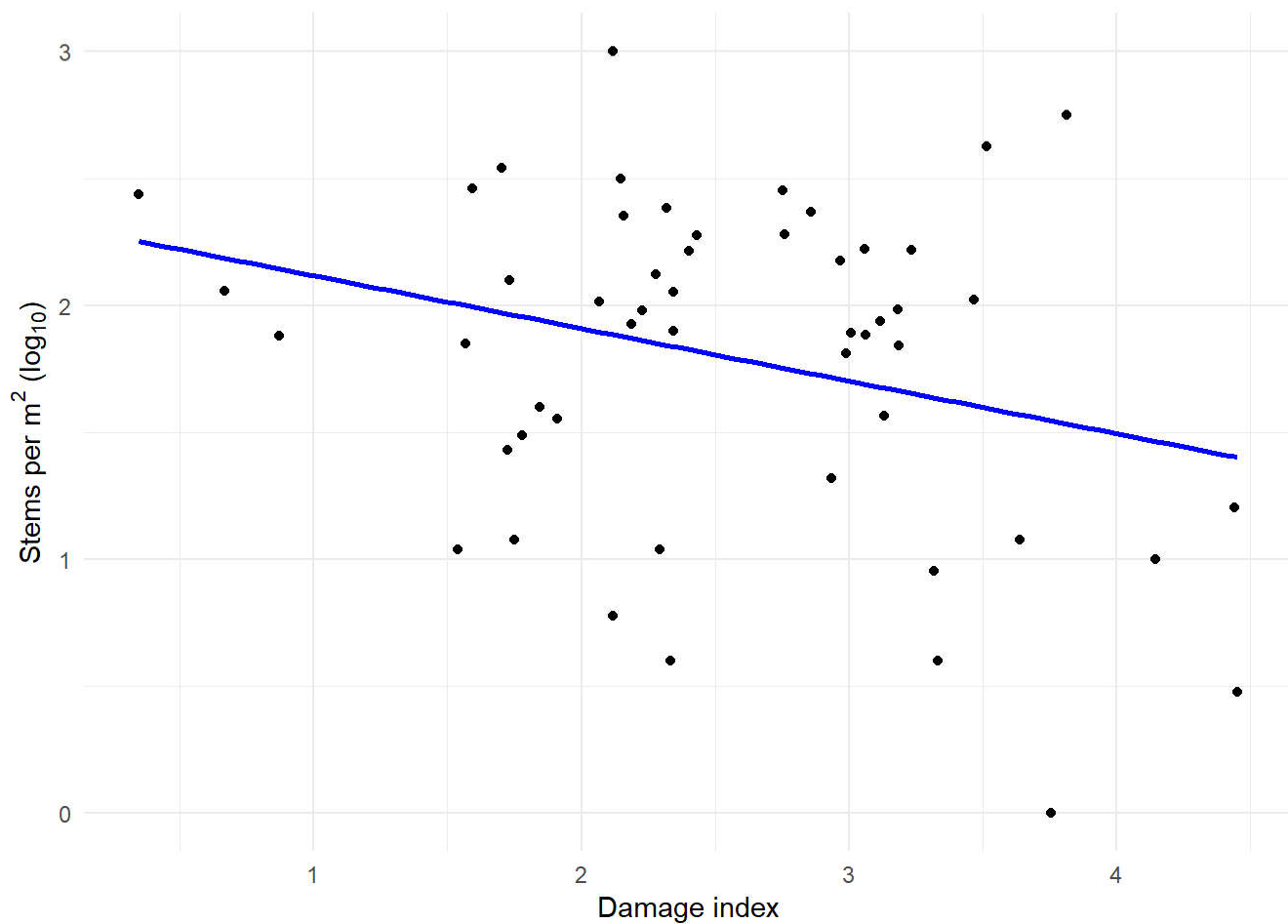
Figure 3. Stem density (log transformed) plotted against damage index. The blue line is a linear regression representing the effect of damage on stem density.

   10. Code that uses linear regression to evaluate H3.

```
dmod <- lm(logDense~damage, loosDat)
summary(dmod)
```

```
## 
## Call:
## lm(formula = logDense ~ damage, data = loosDat)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.5472  -0.4131   0.1259   0.4652   1.2173 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   2.32693    0.26785   8.687 1.47e-11 ***
## damage       -0.20754    0.09909  -2.095   0.0413 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.628 on 50 degrees of freedom
## Multiple R-squared:  0.08067,    Adjusted R-squared:  0.06228 
## F-statistic: 4.387 on 1 and 50 DF,  p-value: 0.0413
```

11. Code and a figure testing the normality of residuals assumption for H3.

```
H3n <- ggplot(data.frame(residuals(dmod))) +
  geom_histogram(aes(x = residuals(dmod), y = ..density..), binwidth = 0.2) +
  stat_function(fun= dnorm, args= list(mean= mean(residuals(dmod)),
                                       sd = sd(residuals(dmod))),colour = "blue", lwd
= 2) +
  scale_x_continuous (name ="Residuals") +
  scale_y_continuous (name ="Density") +
  theme_minimal()
 shapiro.test(residuals(dmod))
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  residuals(dmod)
## W = 0.96791, p-value = 0.1722
```

12. Code and a figure testing the independence of residuals assumption for H3.

```
loos.dmod <- loosDat %>% mutate(resids = residuals(dmod), pred = fitted(dmod))
H3i <- ggplot(loos.dmod, aes(x = pred, y = abs(resids))) +
  geom_point(size = 3) +
  geom_smooth(method = "lm", se = F) +
  scale_x_continuous (name ="Predicted residuals") +
  scale_y_continuous (name ="Residuals") +
  theme_minimal()

summary(lm(abs(resids)~pred, data = loos.dmod))
```

```
##
## Call:
## lm(formula = abs(resids) ~ pred, data = loos.dmod)
##
## Residuals:
##      Min       1Q    Median       3Q      Max
## -0.49098 -0.26037 -0.06974  0.15070  0.92314
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.3634     0.4779   2.853  0.00628 **
## pred         -0.4778     0.2646  -1.805  0.07703 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3481 on 50 degrees of freedom
## Multiple R-squared:  0.0612, Adjusted R-squared:  0.04243
## F-statistic:  3.26 on 1 and 50 DF,  p-value: 0.07703
```

```
fourfigure <- ggarrange(H3n,H3i,ncol = 2, nrow = 1, labels = c("A","B"))
fourfigure
```
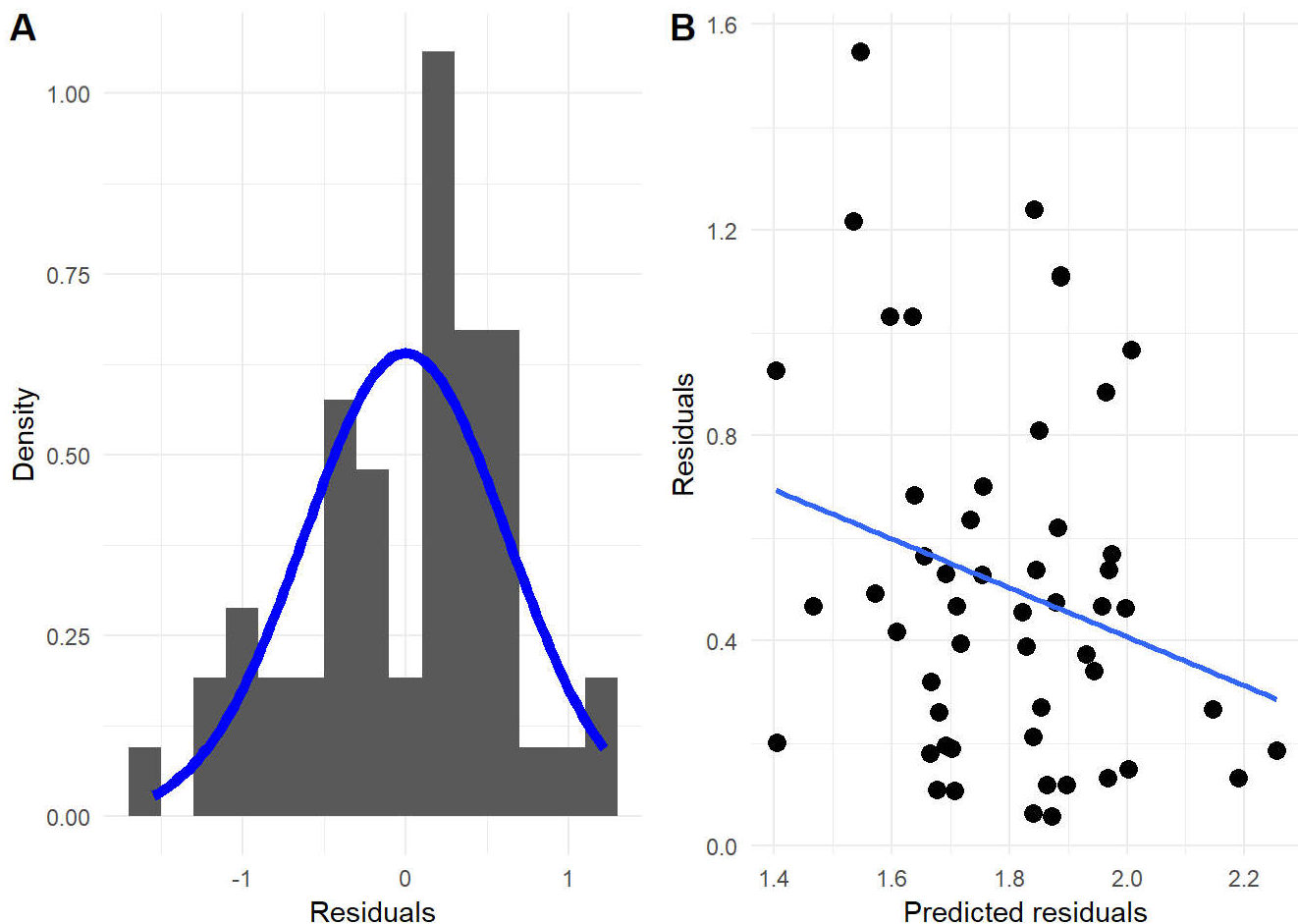

Figure 4. Assessment of the assumptions that residuals of the multivariate data between plant density and

beetle damage was A) normally distributed, and B) homoscedastic.

13. A permutation test for H3.

```
sum.dmod <- summary(dmod)
F.obs <- as.numeric(sum.dmod$fstatistic[1] )

Frand <- numeric(nreps)
for (i in 1:nreps){
  rDense <- sample(loosDat$logDense, length(loosDat$logDense), replace = F)
  sumrmod <- summary(rmod <- lm(rDense~damage, data = loosDat))
  Frand[i] <- as.numeric(sumrmod$fstatistic[1])


}
mean(Frand >= F.obs)
```

```
## [1] 0.04
```

14. A section of text interpreting your results with respect to H3. *(H3) Damage by beetles reduced the local abundance of L. salicaria*

Damage from Galerucella beetles has a significant negative assoiation with plant density. This can be seen in the negative regression line that appears on Figure 4, and in the significant (p = 0.0413) f-statistic derived from evaluating the linear regression of density against damage.

When analyzing the assumptions of normality and homoscedasticisity of residuals, that are required for glm analysis, it is determined that we cannot reject the null hypothesis that the residuals are normal (p = 0.1722) or independant (p = 0.07703).

```
The data supports H3, with glm analysis revealing damage from Galerucella results in
a significant decrease in stem density, a measure of local abundance.
```

15. All figures must have "professional" figure captions. If you are still wondering how to write a proper figure caption please look at the captions of figures in published scientific papers.

16. All figures must be super-pretty, publication-ready and finessed using the scale…() and theme() functions in ggplot.

17. The figures that check linear model assumptions for each hypothesis should be put together in a composite figure using functions in the cowplot package (i.e. 2 composite figures: one for H2 and another for H3).

18. All formatting conventions used in assignment #3 apply here.

19. Make sure all code is as concise as possible and neatly organized in code chunks.