

# Final Paper Code

Jack Magee

2022-12-01

Required packages

```
#install and load dismo  
#install.packages("dismo")
```

```
#read in the data (Level 0) from GitHub
```

```
deer_data <- read.csv('https://raw.githubusercontent.com/mageejac/Eco-870-Final/main/DBS_2022_raw.csv')
```

Data Cleaning

```
#remove duplicated observations of the same group
```

```
deer_data_2 <- deer_data[!duplicated(deer_data$Group.GID), ]
```

```
#subset the data for group size, habitat, and location
```

```
deer_data_3 <- deer_data_2[,c("habitat_group", "total_deer_group", "obs.x", "obs.y")]
```

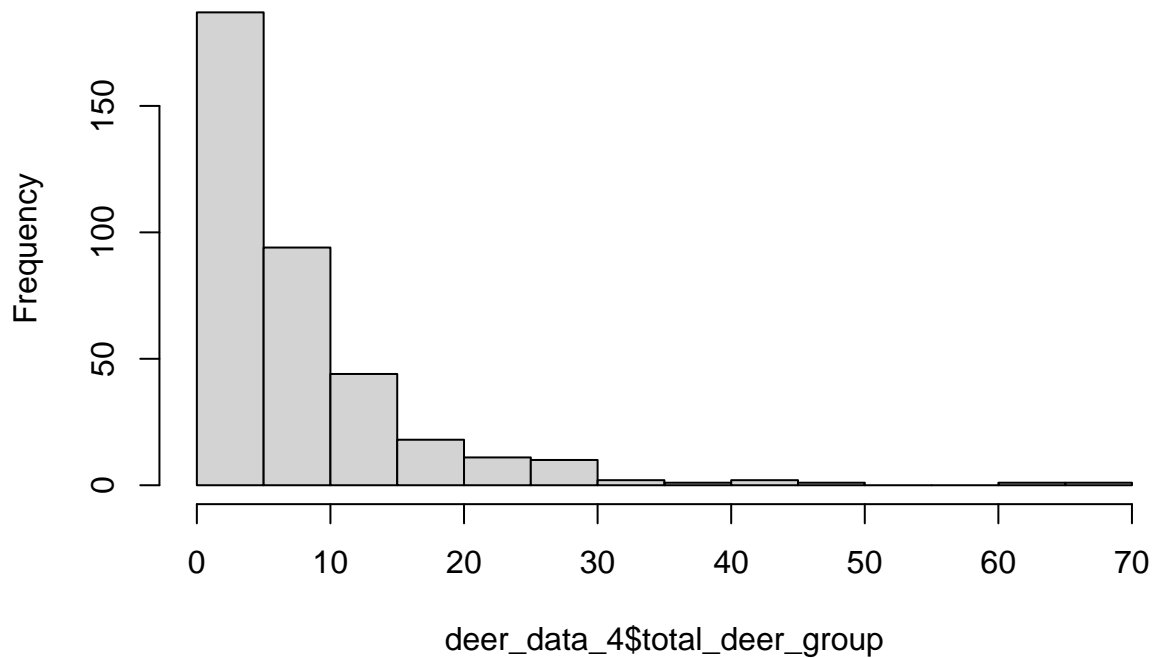
```
#remove observations with unknown habitat type
```

```
deer_data_4 <- deer_data_3[!(deer_data_3$habitat_group=="unk" | deer_data_3$habitat_group=="tilled_unk")]
```

```
# plot data distribution to choose distribution
```

```
hist(deer_data_4$total_deer_group)
```

## Histogram of deer\_data\_4\$total\_deer\_group



```
#choose negative binomial distribution because of large different between variance and mean and over di  
mean(deer_data_4$total_deer_group)
```

```
## [1] 8.153226
```

```
var(deer_data_4$total_deer_group)
```

```
## [1] 72.91716
```

```
Modeling th data
```

```
#load packages
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.2.2
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.2.2
```

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

```
#relevel data with residential as a control
```

```
deer_data_4$habitat_group = as.factor(deer_data_4$habitat_group)
```

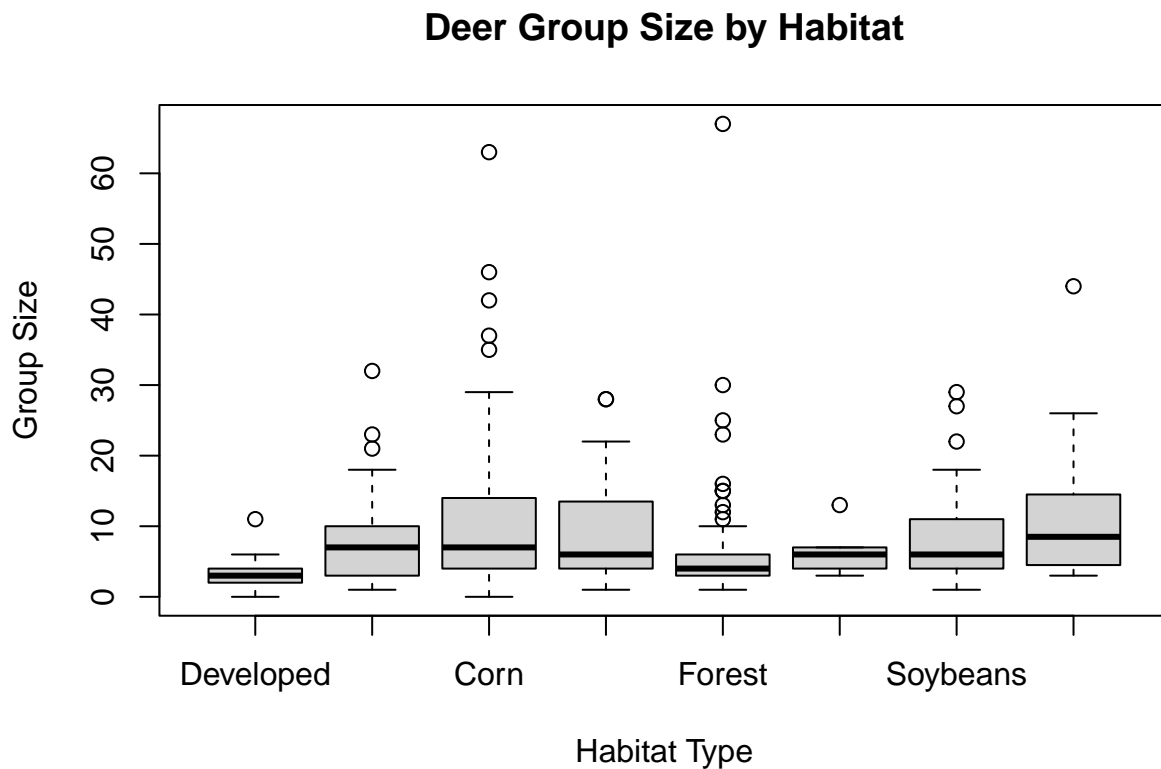
```
deer_data_4$habitat_group_ref <- relevel(deer_data_4$habitat_group, ref = "Developed")
```

```
# create glm
```

```
model_1 <- glm.nb(total_deer_group ~ habitat_group_ref, data = deer_data_4)
```

```
#generating summary statitics
```

```
boxplot(total_deer_group ~ habitat_group_ref, data = deer_data_4,
        main = "Deer Group Size by Habitat",
        xlab = "Habitat Type",
        ylab = "Group Size"
)
```



```
Anova(model_1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: total_deer_group
##          LR Chisq Df Pr(>Chisq)
## habitat_group_ref  50.576  7  1.113e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model_1)
```

```
##
## Call:
## glm.nb(formula = total_deer_group ~ habitat_group_ref, data = deer_data_4,
##        init.theta = 1.837655393, link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```

```

## -2.6332 -0.9686 -0.4332 0.3060 4.8434
##
## Coefficients:
##
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.1676 0.1748 6.681 2.38e-11 ***
## habitat_group_refAlfalfa 0.8486 0.2093 4.054 5.04e-05 ***
## habitat_group_refCorn 1.1632 0.1896 6.136 8.47e-10 ***
## habitat_group_refFallow/Idle Cropland 1.0260 0.2274 4.513 6.40e-06 ***
## habitat_group_refForest 0.6500 0.1962 3.314 0.000921 ***
## habitat_group_refShrubland 0.7195 0.4119 1.747 0.080709 .
## habitat_group_refSoybeans 1.0023 0.2247 4.460 8.18e-06 ***
## habitat_group_refWinter Wheat 1.2962 0.2643 4.904 9.40e-07 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.8377) family taken to be 1)
##
## Null deviance: 431.68 on 371 degrees of freedom
## Residual deviance: 381.11 on 364 degrees of freedom
## AIC: 2288
##
## Number of Fisher Scoring iterations: 1
##
##
## Theta: 1.838
## Std. Err.: 0.159
##
## 2 x log-likelihood: -2269.959

```