**###**

**#R SCRIPT FOR MINI PROJECT#**

**#Packages used**

rm(list=ls())

library(ggplot2)

library(dplyr)

library(tidyr)

library(MuMIn)

library(lme4)

library(blmeco)

library(DHARMa)

library(lmtest)

**#data was downloaded from** *(https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/animal-disease-information/avian/avian-influenza/hpai-2022/2022-hpai-wild-birds)*

**#First into excel as text avoid formatting issues particularly with the date column**

**#Some spelling mistakes in species and sampling method were detected and corrected**

**#Import data from excel**

HPAI\_data <- read.csv("DATA\_FOR\_PROJECT.csv", header = TRUE)

**#Remove data from 2022 so detections are only for 2022**

HPAI\_data\_2022 <- subset(HPAI\_data, Year!="2023")

**#Removed detections with Agency in sampling coloumn as unclear from metadata what this meant**

HPAI\_data\_2022\_Samp <- subset(HPAI\_data\_2022, Sampling!="Agency")

**#Look at data**

View(HPAI\_data\_2022\_Samp)

**#Remove data from captive birds as these were not free in the environment the same way as wild birds**

HPAI\_data\_2022\_Samp\_WOAH <- subset(HPAI\_data\_2022\_Samp, WOAH.Classification!="Captive wild bird")

**#Create vector for the species targeted in the APHIS live bird surveillance plan**

target\_sp<- c("American Black Duck", "American green-winged Teal", "American wigeon",

"Blue-winged teal", "Cinnamon teal", "Fulvous whistling duck", "Gadwall",

"Mallard","Mottled duck", "Northern pintail", "Northern shoveler", "Wood duck")

**#Assign to each whether detection was a Target or Non target species**

HPAI\_data\_2022\_Samp\_WOAH$Target <- ifelse(HPAI\_data\_2022\_Samp\_WOAH$Bird.Species%in%target\_sp,

"Target", "Nontarget")

**#Summarize data to obtain count by State,sampling, month and target#**

sum\_data <- HPAI\_data\_2022\_Samp\_WOAH%>%group\_by(State, Month, Target, Sampling)%>%summarise(Count=n())

**#Remove detection counts that contains dead target species and live non-target samples to allow to test hypothesis**

sum\_data2 <- subset(sum\_data, (sum\_data$Sampling == "Morbidity/Mortality" & sum\_data$Target == "Nontarget")

| (sum\_data$Sampling == "Live bird" & sum\_data$Target == "Target"))

**#Now can summarize data by State, month and whether the counts are from target (live) or nontarget (dead)**

sum\_data3 <- sum\_data2%>%group\_by(State, Month, Target)%>%summarise(Count=sum(Count))

**#Seperate into count columns for target and nontarget**

sum\_data4 <- sum\_data3%>%pivot\_wider(names\_from = "Target", values\_from = "Count")

**#Remove any rows with NAs as will not be fitted in model**

sum\_data5 <- na.omit(sum\_data4)

**#ensure month and state and factor**

sum\_data5$State <- as.factor(sum\_data5$State)

sum\_data5$Month <- as.factor(sum\_data5$Month)

**#Test that mean and variance in response is similar**

var(sum\_data5$Target)

mean(sum\_data5$Target)

**#Mean and var are different try logTarget - now more similar**

sum\_data5$logTarget <- log(sum\_data5$Target)

mean(sum\_data5$logTarget)

var(sum\_data5$logTarget)

**#Look at data**

str(sum\_data5)

View(sum\_data5)

**#Modelling - count data with random effects due to state and month**

**#Use Poisson GLMM**

model1 <- glmer(Nontarget ~ logTarget + (1|Month) + (1|State), data = sum\_data5, family = "poisson")

model2 <- glmer(Nontarget ~ logTarget + (1|Month) + (1|State) + (1|Month:State), data = sum\_data5, family = "poisson")

model3 <- glmer(Nontarget ~ logTarget + (1|Month:State), data = sum\_data5, family = "poisson")

summary(model1)

summary(model2)

summary(model3)

**#Check dispersion - 2 and 3 are closest to 1**

dispersion\_glmer(model1)

dispersion\_glmer(model2)

dispersion\_glmer(model3)

**#Check rsquared - similar rsquared value (0.87) more explained by random effects in model 3**

r.squaredGLMM(model2)

r.squaredGLMM(model3)

**#Look at goodness of fit on plots no overall deviations from both**

m2\_resi <- simulateResiduals(model2)

m3\_resi <- simulateResiduals(model3)

plot(m2\_resi)

plot(m3\_resi)

**#Compare models using likelihood Ratio Test - model 3 has statistically better fit**

lrtest(model3, model2)

**#MODEL 3 USED#**

summary(model3)

**###**

#Not for report#

#Intially looking at how detection in target species related to Mortality in target species

#Keep Target deaths instead

sum\_data2.1 <- subset(sum\_data, (sum\_data$Sampling == "Morbidity/Mortality" & sum\_data$Target == "Target")

| (sum\_data$Sampling == "Live bird" & sum\_data$Target == "Target"))

sum\_data3.1 <- sum\_data2.1%>%group\_by(State, Month, Sampling)%>%summarise(Count=sum(Count))

sum\_data4.1 <- sum\_data3.1%>%pivot\_wider(names\_from = "Sampling", values\_from = "Count")

sum\_data5.1 <- na.omit(sum\_data4.1)

#Model

model2.1 <- glmer(sum\_data5.1$`Morbidity/Mortality` ~ sum\_data5.1$`Live bird` + (1|Month:State),

data = sum\_data5.1, family = "poisson")

#Close to significant effect as expected

summary(model2.1)

#Rsquared close to 1 - >0.9

dispersion\_glmer(model2.1)

r.squaredGLMM(model2.1)

#Scatterplot

p1 <- ggplot(sum\_data5.1, aes(x = log(`Live bird`), y = log(`Morbidity/Mortality`)))

p1 <- p1 + geom\_point(size = 2)

p2 <- p1 + geom\_smooth(method = "lm", se = TRUE)

p3 <- p2 + labs(x="Live detection", y = "Mortality")

p3

###