

```

load("ElkData.Rdata")
##pairs(ElkData)
library(car)
rownames(ElkData) = 1:nrow(ElkData)
#####

## Initial Poisson model fitting
fit = glm(num.seen~season+elev+perc.forest+perc.open.water+perc.developed+obs.days,family=poisson,data=ElkData)
summary(fit)
par(mfrow=c(2,2))
plot(fit)
crPlots(fit)
outlierTest(fit,n.max=1000)
vif(fit)
#####

## Model fitting with only significant predictors
fit = glm(num.seen~season+elev+perc.open.water+perc.developed,family=poisson,data=ElkData)
summary(fit)
par(mfrow=c(2,2))
plot(fit)
crPlots(fit)
outlierTest(fit,n.max=1000)
vif(fit)
#####

## All outlier indices were indentified and written to a .csv file
outliers = read.csv("outliers.csv")
idx = outliers$Idx
## The reduced dataset is used for future modeling
newdata=ElkData[-idx,]
#####

fit.new = glm(num.seen~season+elev+perc.open.water+perc.developed,family=poisson,data=newdata[-c(4,7)])
summary(fit.new)
outlierTest(fit.new,n.max=1000)
crPlots(fit.new)
vif(fit.new)
plot(fit.new)
#####

## Significant predictors were identified and included in the model
fit.best = glm(num.seen~season+elev+perc.open.water+perc.developed+elev*perc.developed,family=poisson,data=newdata)
summary(fit.best)
outlierTest(fit.best,n.max=1000)
crPlots(fit.best)
vif(fit.best)
plot(fit.best)
#####

## simulation study to see if resids are acceptable
## simulate data from the fitted model
muhat=predict(fit.best,newdata=newdata,type="response")
ysim=rpois(length(muhat),lambda=muhat)
fit.sim=glm(ysim~season+elev+perc.open.water+perc.developed+elev*perc.developed,family="poisson",data=newdata)
summary(fit.sim)
plot(fit.sim)
## The simulated residuals looks similar in quality to the ones derived from the actual data
## Conclusion: No particular problems with the residuals. Hence, model can be considered to be ok.
#####

##Plotting for final model
coef = fit.best$coefficients
par(mfrow=c(3,2))
#season(P)
X = seq(0,1,1/1000)
Y = (exp(coef[1]*X))
plot(X,Y,xlab = "Season:P",ylab = "Mean count of Elk observed")
#season(SF)
X = seq(0,1,1/1000)
Y = (exp(coef[2]*X))
plot(X,Y,xlab = "Season:SF",ylab = "Mean count of Elk observed")
#season(W)
X = seq(0,1,1/1000)
Y = (exp(coef[3]*X))
plot(X,Y,xlab = "Season:W",ylab = "Mean count of Elk observed")
#Elev
X = seq(min(newdata$elev),max(newdata$elev), (max(newdata$elev)-min(newdata$elev))/1000)
Y = (exp(coef[4]*X))
plot(X,Y,xlab = "Elev",ylab = "Mean count of Elk observed")
#perc.open.water
X = seq(min(newdata$perc.open.water),max(newdata$perc.open.water), (max(newdata$perc.open.water)-min(newdata$perc.open.water))/1000)
Y = (exp(coef[5]*X))
plot(X,Y,xlab = "Perc.open.water",ylab = "Mean count of Elk observed")
#perc.developed

```

```

X = seq(min(newdata$perc.developed),max(newdata$perc.developed), (max(newdata$perc.developed)-
min(newdata$perc.developed))/1000)
Y = (exp(coef[6]*X))
plot(X,Y,xlab = "Perc.developed",ylab = "Mean count of Elk observed")
#####

fit.mod = step(glm(num.seen~.^2,family=poisson,data=newdata[, -c(2)]))
summary(fit.mod)

#Season:P
newdata2 = newdata[newdata$season == 'P',]
muhat.pred.P=predict(fit.mod,newdata=newdata2[, -c(2)],type="response")
ysim=rpois(length(muhat.pred.P),lambda=muhat.pred.P)
mspe.P = mean((newdata2$num.seen-ysim)^2)

#Season:SF
newdata3 = newdata[newdata$season == 'SF',]
muhat.pred.SF=predict(fit.mod,newdata=newdata3,type="response")
ysim=rpois(length(muhat.pred.SF),lambda=muhat.pred.SF)
mspe.SF=mean((newdata3$num.seen-ysim)^2)
plot(fit.mod)

#SeasonW
newdata4 = newdata[newdata$season == 'W',]
muhat.pred.W=predict(fit.mod,newdata=newdata4[, -c(2)],type="response")
ysim=rpois(length(muhat.pred.W),lambda=muhat.pred.W)
mspe.W=mean((newdata4$num.seen-ysim)^2)

#Comparison Table
T=matrix(c(mspe.P,mspe.SF,mspe.W),ncol=1)
rownames(T)=c("Season:P","Season:SF","Season:W")
colnames(T)=c("MSPE")
T
#####

elev=c(-0.954,0.209)
perc.forest=c(0,0.222)
perc.open.water=c(0,0.01)
perc.developed=c(0,0)

pred.data=data.frame(elev,perc.forest,perc.open.water,perc.developed)

pred.data$muhat=predict(fit.mod,newdata=pred.data,type="response")
pred.data$muhat
y.site1=rpois(10000,lambda=pred.data$muhat[1])
y.site2=rpois(10000,lambda=pred.data$muhat[2])
mean(y.site1)
mean(y.site2)
## Based on the average of 10,000 simulated Poisson values,
##Site 1 always has more elk sighting and hence is the better option than Site 2
#####

```