Jamie Burke
ecostats_hw2.R

```
######## Q1 seedling survival #######

seedling<-read.csv("SEEDLING_SURVIVAL.csv")
summary(seedling)

##      survival            HEIGHT           LIGHT
##   Min.   :0.0000   Min.   :  3.50   Min.   : 0.90
##   1st Qu.:1.0000   1st Qu.: 11.00   1st Qu.: 7.64
##   Median :1.0000   Median : 13.80   Median :10.33
##   Mean   :0.8725   Mean   : 18.33   Mean   :10.76
##   3rd Qu.:1.0000   3rd Qu.: 19.80   3rd Qu.:13.38
##   Max.   :1.0000   Max.   :180.00   Max.   :27.80

str(seedling)

## 'data.frame':    1435 obs. of  3 variables:
##  $ survival: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ HEIGHT  : num  47 70.2 16.3 23.5 23 21 30 17.5 76 57.5 ...
##  $ LIGHT   : num  2.4 14.83 9.15 8.62 4.26 ...

####effect of height####


seedlingmod<-glm(seedling$survival~seedling$HEIGHT,family="binomial")

coef(seedlingmod)

##     (Intercept) seedling$HEIGHT
##     -0.06271111      0.14071141

confint(seedlingmod)

## Waiting for profiling to be done...

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred



##                      2.5 %     97.5 %
## (Intercept)     -0.5791061 0.4268167
## seedling$HEIGHT  0.1038803 0.1815477

plogis(-0.062711)

## [1] 0.4843274
```
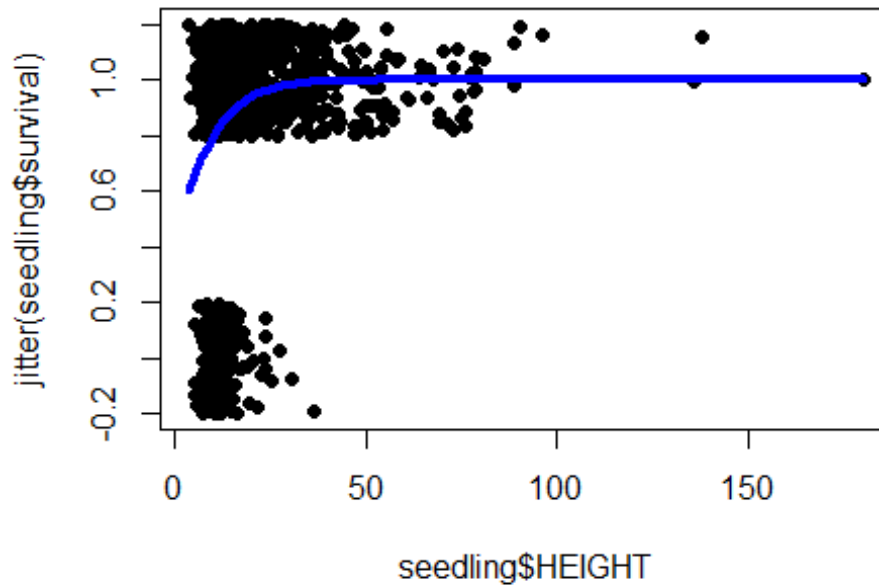
```
plot((jitter(seedling$survival)~seedling$HEIGHT),pch=19)

curve(plogis(-0.06271111+0.14071141*x),add=T,col="blue",lwd=4)
```


The baseline for seedling survival at height zero is 0.484 or about 48.4%. As you increase seed height by 1 unit, seedling survival increases by about 3.5%.

```
#####effect of light#####


seedlingmod2<-glm(seedling$survival~seedling$LIGHT,family="binomial")
coef(seedlingmod2)

##    (Intercept) seedling$LIGHT
##     2.66194692    -0.06552684

confint(seedlingmod2)

## Waiting for profiling to be done...

##                     2.5 %      97.5 %
## (Intercept)     2.25136434   3.0876309
## seedling$LIGHT -0.09841747  -0.0325795

plot(seedling$survival~seedling$LIGHT,pch=1)
curve(plogis(2.66194692+-0.06552684*x),add=T,col="green",lwd=4)
```
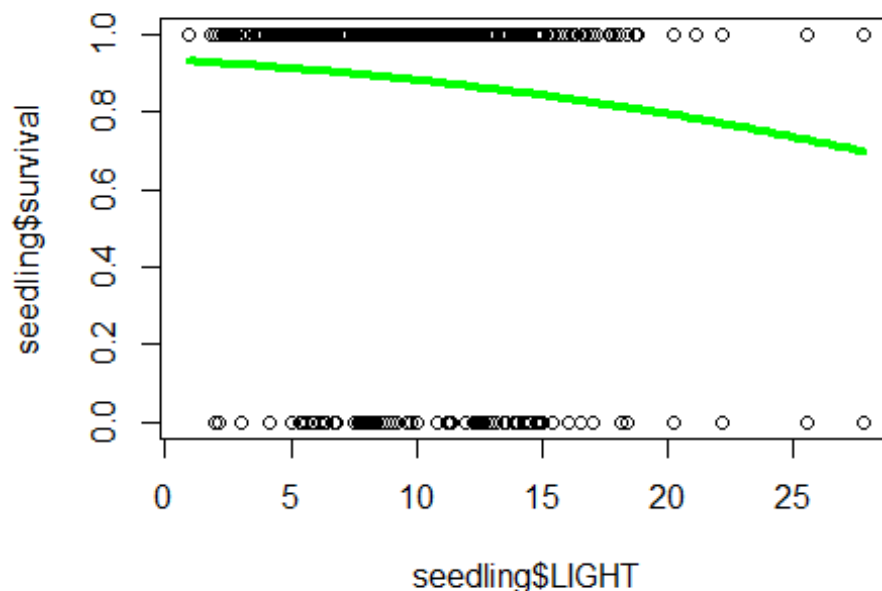
A baseline of seedling survival at light zero is 0.934 or about 93.4% survival which doesn't exactly make sense, but we need to note that our data does not extend to zero light. As light increase by 1 unit, the effect of survival decreases by 0.016 or 1.6%.

The effect size for light is smaller than the effect size for height so I think height is a slightly better predictor for seedling survival although both predictors are significant.

```
#####Q2 seed addition #######

seeds<-read.csv("SEEDS.csv")
str(seeds)

## 'data.frame':    281 obs. of  8 variables:
##  $ Site     : Factor w/ 94 levels "m1","m10","m11",..: 1 1 1 2 2 2 3 3 3 4
...
##  $ Pile     : Factor w/ 281 levels "m1.15","m1.45",..: 1 2 3 4 5 6 7 8 9
10 ...
##  $ DBH      : num  21.6 0 47 0 27.7 ...
##  $ seedlings: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ seeds    : int  15 45 5 15 45 5 15 45 5 15 ...
##  $ recruits : int  2 2 1 0 0 2 6 0 1 1 ...
##  $ grass    : int  1 0 0 0 0 0 0 1 1 0 ...
##  $ light    : num  9.35 17 6.68 6.72 4.91 ...

head(seeds)
```

```
##   Site   Pile     DBH seedlings seeds recruits grass light
## 1   m1  m1.15 21.590        0    15        2     1  9.35
## 2   m1  m1.45  0.000        0    45        2     0 17.00
## 3   m1   m1.5 46.990        0     5        1     0  6.68
## 4  m10 m10.15  0.000        0    15        0     0  6.72
## 5  m10 m10.45 27.686        0    45        0     0  4.91
## 6  m10  m10.5  0.000        0     5        2     0  3.07
```
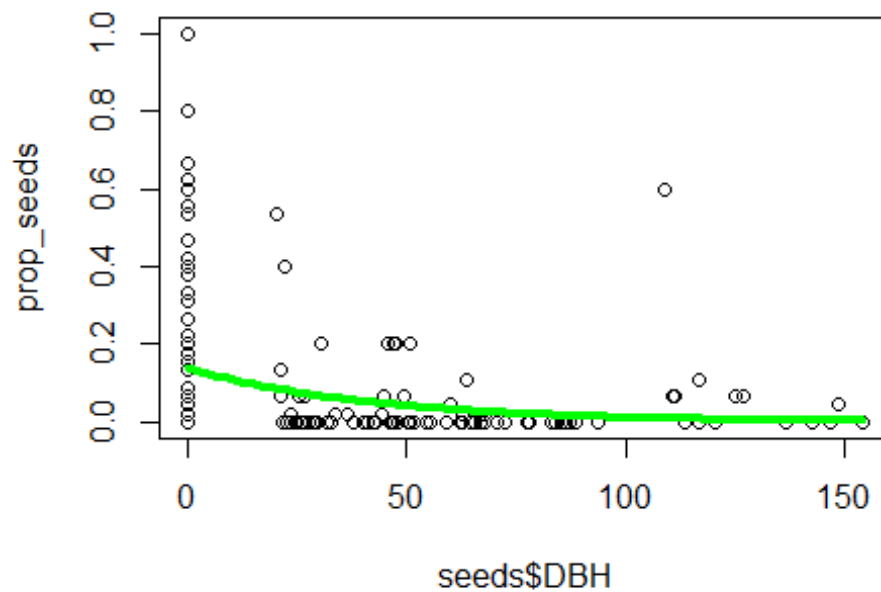
```r
prop_seeds<-seeds$recruits/seeds$seeds
plot(prop_seeds~seeds$DBH)

seed_success<-cbind(seeds$recruits,seeds$seeds-seeds$recruits)
seedsmodel<-glm(seed_success~seeds$DBH,family="binomial")
coef(seedsmodel)
```

```
## (Intercept)    seeds$DBH
## -1.82983245 -0.02625002
```

```r
curve(plogis(-1.8298+-0.02625001*x),add=T,col="green",lwd=4)
```



```r
confint(seedsmodel)
```

```
## Waiting for profiling to be done...
```

```
##                    2.5 %      97.5 %
## (Intercept) -1.92090938 -1.74065050
## seeds$DBH   -0.03105183 -0.02184552
```

```r
plogis(-1.8298)
```

```
## [1] 0.1382621
```

With a 95% CI that does not overlap zero, the results do show that DBH has a significant effect on seedling germination.

Q3 mosquitos

```r
mosquito<-read.csv("mosquito_data.csv")
head(mosquito)
```

```
##   Emergent_adults Egg_Count Detritus
## 1               2         3     0.00
## 2               0         2     0.01
## 3               2         3     0.01
## 4               4         6     0.02
## 5               1         4     0.02
## 6               5         7     0.03
```

```r
emerge_success<-cbind(mosquito$Emergent_adults,mosquito$Egg_Count-
mosquito$Emergent_adults)
mosquitomod<-glm(emerge_success~mosquito$Detritus,family="binomial")

coef(mosquitomod)
```

```
##       (Intercept) mosquito$Detritus
##         1.3240425        -0.3216083
```

```r
confint(mosquitomod)
```

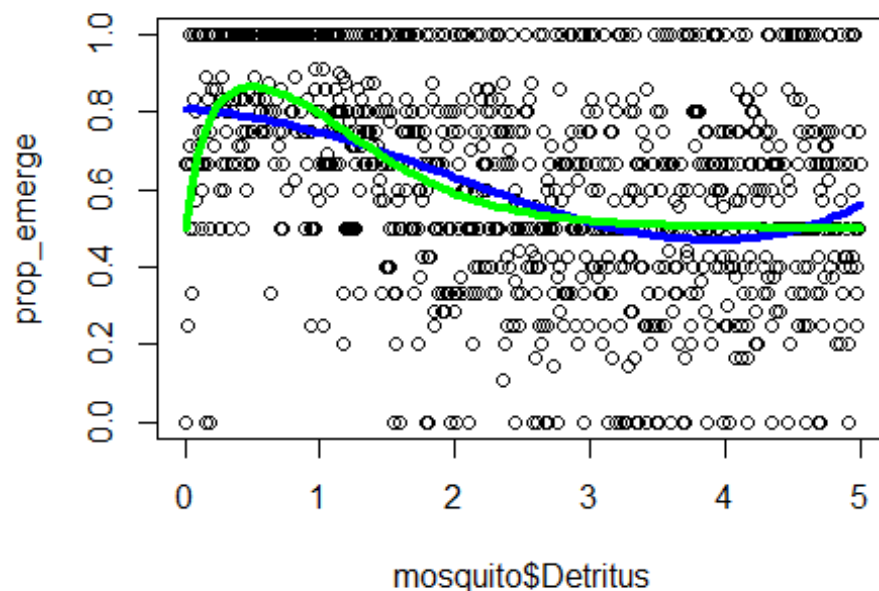```
Waiting for profiling to be done...
```

```
##                       2.5 %     97.5 %
## (Intercept)        1.198777  1.4509989
## mosquito$Detritus -0.364044 -0.2795097
```

```r
prop_emerge<-mosquito$Emergent_adults/mosquito$Egg_Count
plot(prop_emerge~mosquito$Detritus)


curve(plogis(1.44-0.19*x-0.21*x^2+0.04*x^3),add=T,col="blue",lwd=4)

curve(plogis(10*x*exp(-2*x)),add=T,col="green",lwd=4)
```

```
#polynomial#
-
sum(dbinom(x=mosquito$Emergent_adults,size=mosquito$Egg_Count,prob=plogis(1.4
4-0.19*mosquito$Detritus-
0.21*mosquito$Detritus^2+0.04*mosquito$Detritus^3),log=T))
```

[1] 1415.63

```
#Ricker#
-
sum(dbinom(x=mosquito$Emergent_adults,size=mosquito$Egg_Count,prob=plogis(10*
mosquito$Detritus*exp(-2*mosquito$Detritus)),log=T))
```

[1] 1385.847

The polynomial model has a term that has a higher power than the term in for the Ricker model. In this case this determines the curve to be more gradual than the Ricker model showing the detritus has a smaller effect on emergent adult survival than the Ricker model shows. The Ricker model shows that detritus has a greater effect on emergent adult survival at lower levels of detritus than at higher levels where the effect is more gradual. In this case, the Ricker model is the best model because is predicts the lower neg-log likelihood.

```
######Q4######

######linear: level of water predicts ibis nests #####

slope=2
intercept=20
sd=1.5

water_level<-seq(from=0,to=10,length=100)

ibisnests=rnorm(n=100,mean=intercept+slope*water_level,sd)
hist(ibisnests)
```
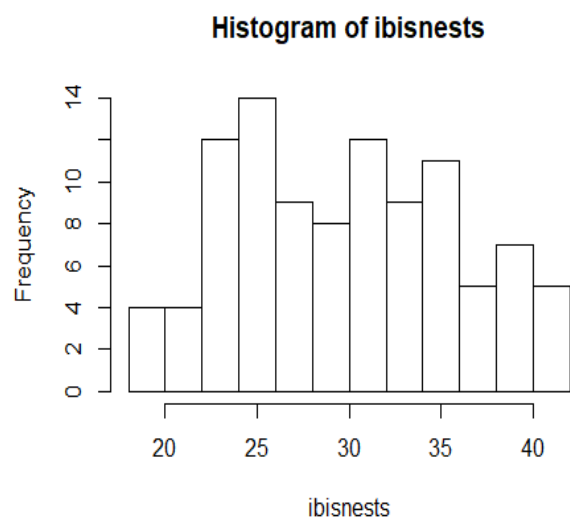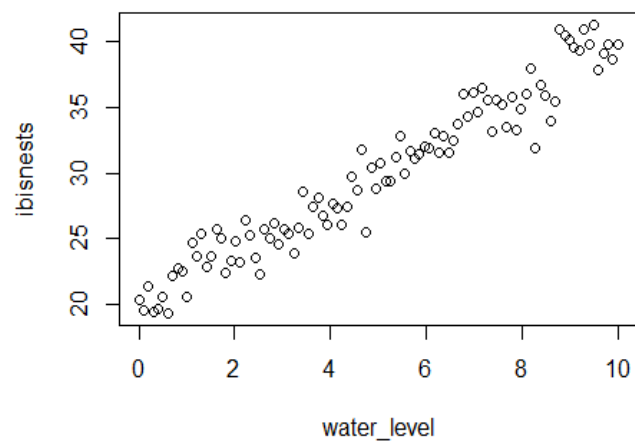

Histogram of ibisnests

```
plot(ibisnests~water_level)
```
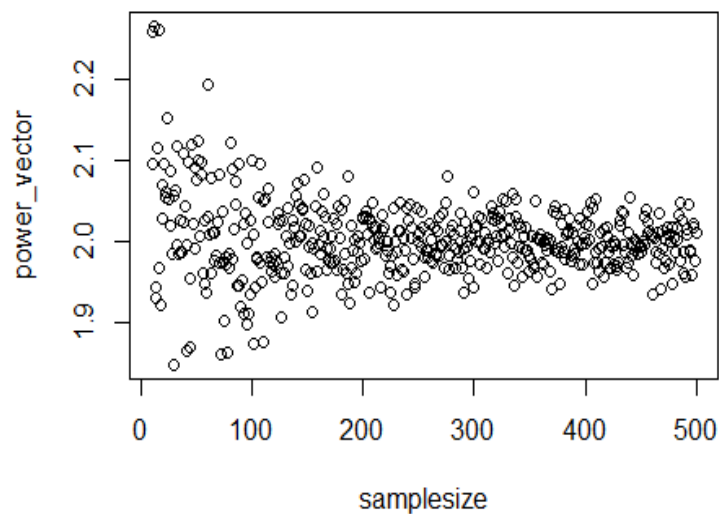
```
samplesize=seq(from=10,to=500)
power_vector<-rep(NA,times=length(samplesize))

for(i in 1:length(samplesize)){
  y=rnorm(n=samplesize[i],
mean=intercept+slope*seq(from=0,to=10,length=samplesize[i]),sd=1.5)
  response=y
  modnests<-glm(response~seq(from=0,to=10,length=samplesize[i]))
  power_vector[i]=coef(modnests)[2]}

plot(power_vector~samplesize)
```



```
summary(modnests)$coefficients[2,4]
[1] 2.057707e-306

mse=function(yhat,y){return(mean((y-yhat)^2))}
> mse(power_vector,2)
[1] 0.002285321
```

This this p-value<0.05 for the slope, we can determine that 500 samples is enough to ensure a significant result for the linear regression. The MSE is also very low and so the estimated slopes are very close to the true slope.

```
#####binomail power analysis######

slope2=0.25
intercept2=0
```
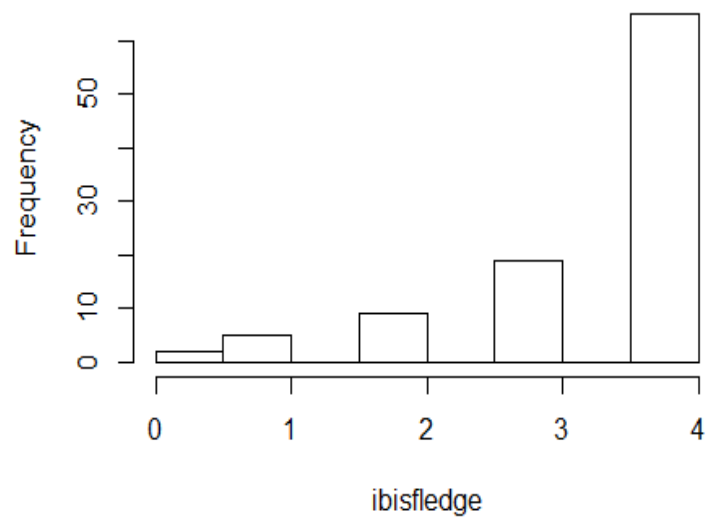
```
forage_time=seq(from=0,to=24,length=100)
samplesize=100
ibisfledge=rbinom(n=samplesize,prob=plogis(intercept2+slope2*seq(from=0,to=24
,length(samplesize))),size=4)
hist(ibisfledge)
```
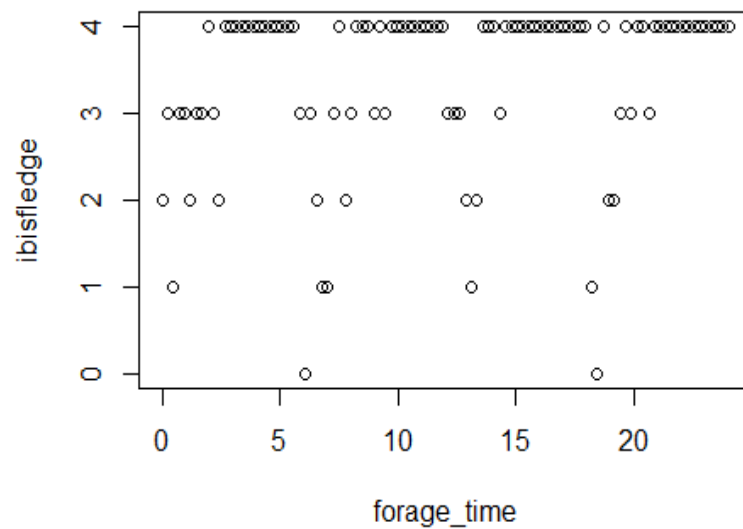
Histogram of ibisfledge


```
plot(ibisfledge~forage_time)
```

```
yibis2=cbind(ibisfledge,4-ibisfledge)
modfledge<-glm(yibis2~forage_time,family="binomial")
coef(modfledge)

## (Intercept) forage_time
##  1.02660348  0.06493152

samplesize2=rep(seq(from=10,to=200),time=20)
forage_time=seq(from=0,to=24,length=100)

power_vector2=rep(NA,times=length(samplesize2))

for(i in 1:length(samplesize2)){

y=rbinom(n=samplesize2[i],prob=plogis(intercept+slope*seq(from=0,to=24,length
=samplesize2[i])),size=4)
  response=cbind(y,4-y)
  modfledge2<-
glm(response~seq(from=0,to=24,length=samplesize2[i]),family="binomial")
  power_vector2[i]=coef(modfledge2)[2]}

## Warning: glm.fit: algorithm did not converge

plot(power_vector2~samplesize2
```
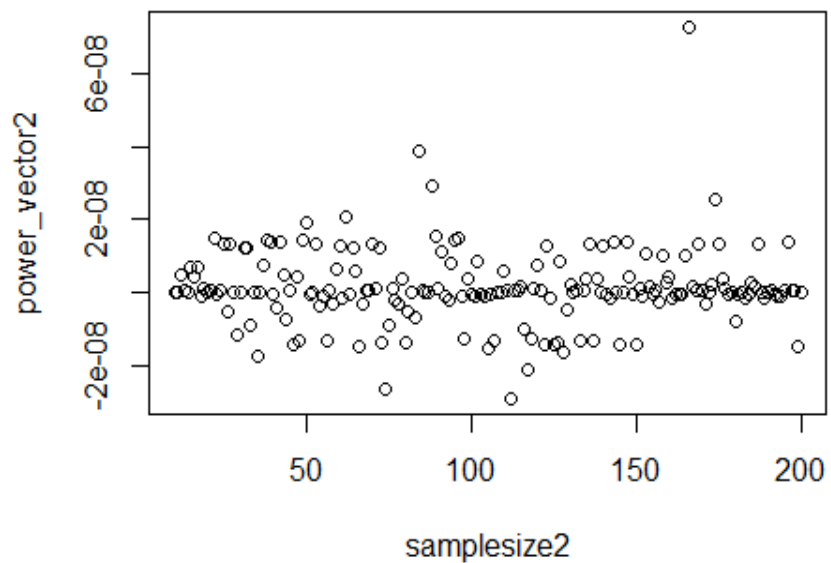
```
summary(modfledge2)$coefficients[2,4]
[1] 1

> mse=function(yhat,y){return(mean((y-yhat)^2))}
> mse(power_vector2,0.25)
[1] 0.0625
```

This this p-value<0.05 for the slope, we can determine that 200 samples is not enough to ensure a significant result for the binomial regression. We would have to run the model again to determine the number of sample sizes needed to get a significant result although it does look like with 200 samples, it is started to estimate the true slope better but you would still need a much larger sample size to get a significant p value. With the MSE of 0.0625, the model does provide a good estimate of the true slope.

In general, statistical power is general higher than with discrete response variable because there can be a larger amount of possible values and this usually will decrease the variation and provides more statistical power.