

## HW\_2

Kristina Parker

1/31/2019

Question 1 a) effect of height on seedling survival

```
seed <- read.csv("SEEDLING_SURVIVAL.csv")
seedmod <- glm(seed$survival ~ seed$HEIGHT, family = "binomial")
coef(seedmod)

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

plogis(-0.06271111)

## [1] 0.4843274

0.14071141 /4

## [1] 0.03517785

confint(seedmod)

## Waiting for profiling to be done...

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

plogis(confint(seedmod))

## Waiting for profiling to be done...

##           2.5 %    97.5 %
## (Intercept) 0.3591383 0.6051133
## seed$HEIGHT 0.5259467 0.5452627

0.1038803/4

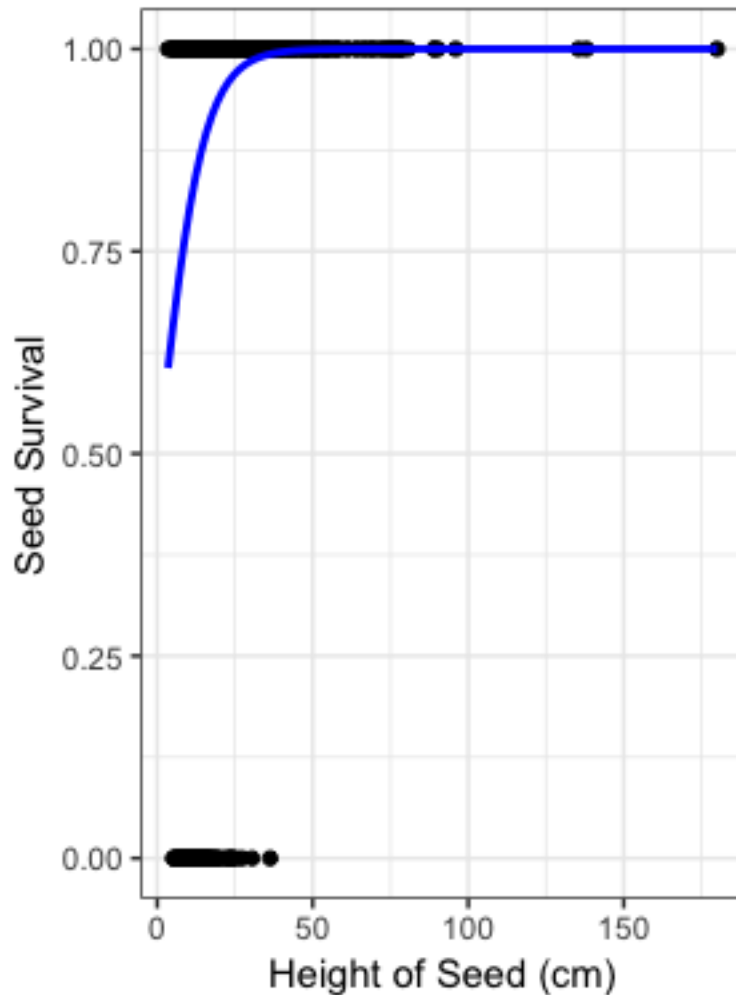
## [1] 0.02597007

0.1815477/4

## [1] 0.04538693

library("ggplot2")
#ggplotgui::ggplot_shiny(dataset = seed)
seedplot <- ggplot(seed, aes(x = HEIGHT, y = survival)) +
  geom_point() +
```

```
labs(x = 'Height of Seed (cm)', y = 'Seed Survival') +
stat_function(fun = function(x) plogis(-0.06271111 + 0.14071141 * x), col =
"blue", lwd = 1) +
theme_bw()
seedplot
```



*A seed survival is 48% (CI 35% - 61%) when the seeds height it zero. An increase in seed height positively effects seed survival by 3.5% (CI 2.5% - 4.5%). Once seeds reached a height greater than 36 cm the seeds always survived.*

b) effect of light on seedlings survival

```
lightmod <- glm(seed$survival ~ seed$LIGHT, family = "binomial")
coef(lightmod)
```

```
## (Intercept) seed$LIGHT
## 2.66194692 -0.06552684
```

```
plogis(2.66194692)
```

```
## [1] 0.9347435
```

```

-0.06552684/4

## [1] -0.01638171

confint(lightmod)

## Waiting for profiling to be done...

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

plogis(confint(lightmod))

## Waiting for profiling to be done...

##              2.5 %      97.5 %
## (Intercept)  0.9047682  0.9563796
## seed$LIGHT   0.4754155  0.4918558

-0.09841747/4

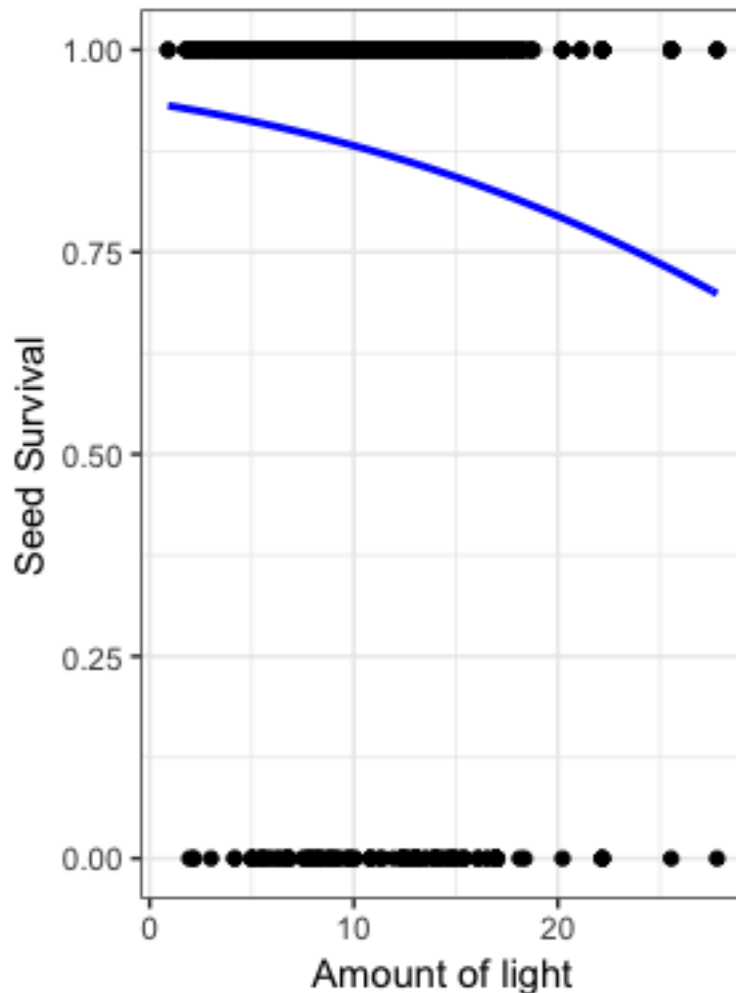
## [1] -0.02460437

-0.0325795/4

## [1] -0.008144875

#ggplotgui::ggplot_shiny(dataset = seed)
lightplot <- ggplot(seed, aes(x = LIGHT, y = survival)) +
  geom_point()+
  labs(x = 'Amount of light', y = 'Seed Survival') +
  stat_function(fun = function(x) plogis(2.66194692 + -0.06552684 * x), col =
"blue", lwd = 1) +
  theme_bw()
lightplot

```



*At very low light seed survival is 93% (CI 90% - 96%). Increasing the amount of light negatively effects seed survival by 1.6% (CI 2.5% - 0.8%).*

Is height or light a stronger predictor of seedling survival?

Height is a stronger predictor of seedling survival. It has a greater effect on seedling survival.

```
seeds <- read.csv("seeds.csv")
prop_seeds <- seeds$recruits/seeds$seeds
seed_response <- cbind(seeds$recruits, seeds$seeds - seeds$recruits)
seedling.glm <- glm(seed_response ~ seeds$seedlings, family = "binomial")
coef(seedling.glm)

##      (Intercept) seeds$seedlings
##      -2.035570      -1.213717

plogis(-2.035570)

## [1] 0.1155186
```

```

-1.213717/4

## [1] -0.3034292

confint(seedling.glm)

## Waiting for profiling to be done...

##              2.5 %      97.5 %
## (Intercept)   -2.121808 -1.9511837
## seeds$seedlings -1.592362 -0.8977661

plogis(confint(seedling.glm))

## Waiting for profiling to be done...

##              2.5 %      97.5 %
## (Intercept)    0.1069952 0.1244243
## seeds$seedlings 0.1690519 0.2895098

-1.592362/4 #CI 2.5%

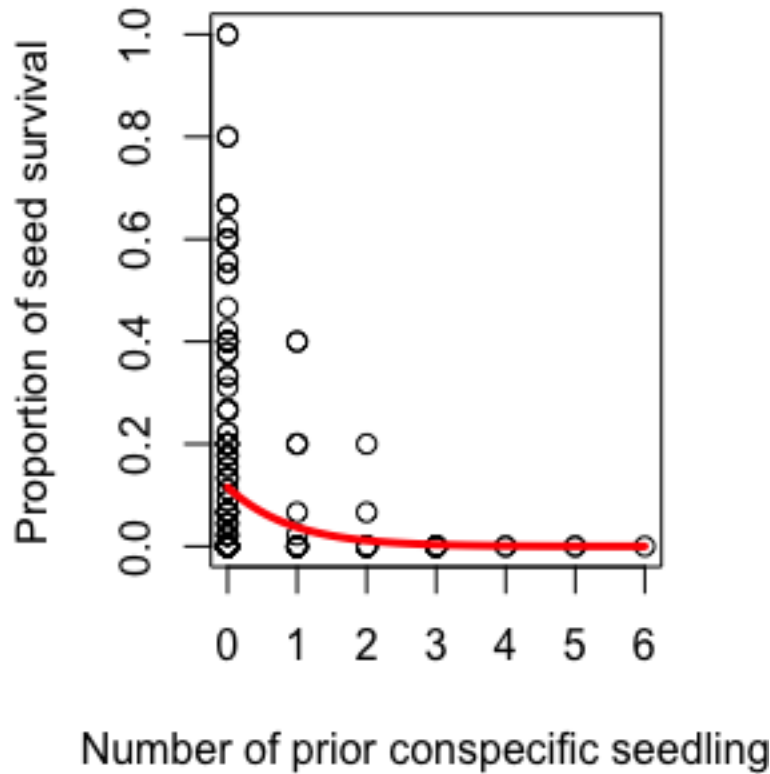
## [1] -0.3980905

-0.8977661/4 # CI 97.5%

## [1] -0.2244415

plot(prop_seeds ~ seeds$seedlings, xlab = "Number of prior conspecific
seedlings", ylab = "Proportion of seed survival")
curve(plogis(-2.035570 + -1.213717 * x), add = T, col = "red", lwd = 3)

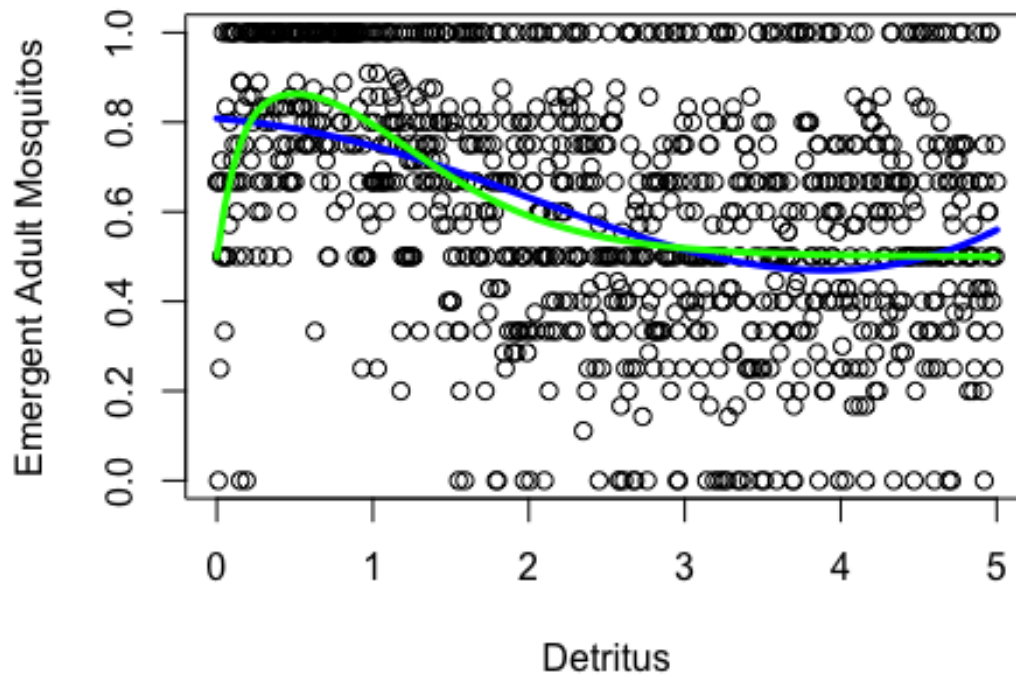
```



Do your results support the hypothesis that your selected predictor variable has a significant effect on seedling germination?

When no prior conspecific seedlings were present at a site 12% (CI 11% - 12%) of the additional seeds successfully germinated. The effect of prior conspecific seedlings reduces the chances of successful germination by 30% (CI 17% - 29%). The predictor variable, prior conspecific seedlings, does have a significant effect on additional seedling germination. This implies that higher number of seedlings reduces the chances of germination, hence implies these seeds follow negative-density dependency population size model. The survival of this seed depends on long-distance seed dispersal.

```
mos <- read.csv("mosquito_data.csv")
prop_mos <- mos$Emergent_adults / mos$Egg_Count
plot(prop_mos ~ mos$Detritus, xlab = "Detritus", ylab = "Emergent Adult Mosquitos")
curve(plogis(1.44 - 0.19*x - 0.21 * x^2 + 0.04 * x^3), add = T, col = "blue", lwd = 3) #polynomial
curve(plogis(10 * x * exp(-2*x)), add = T, col = "green", lwd = 3) # Ricker
```



How are the biological implications of the polynomial model different from the Ricker model?

The polynomial model shows that the proportion of emergent adult mosquitos gradually declines as organic detritus increases in pools of water. Until the level of organic detritus reaches  $\sim 3.7$  then the proportion of emergent adults begins to increase gradually. Biologically this model implies that pools of water with no organic detritus have a high proportion of emergent adult mosquitos from eggs laid. At low levels of detritus increases this negatively effects the proportion of emergent adults; however, at higher levels of detritus has a positive effect on the proportion of emergent adults. The Ricker model shows that the proportion of emergent adult mosquitos rapidly increases with low levels of detritus but as levels increase the proportion of emergent adults declines and plateaus. The biological implications of this model is adult mosquitos emerging from eggs in pools of water is about 50%, this proportion increases with slight levels of organic detritus. As the level of detritus increases the proportion of emerging adults declines back to 50% of mosquitos emerging.

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

```
## [1] 1385.847
```

```
-sum(dbinom(mos$Emergent_adults, prob = plogis(1.44-.19 * mos$Detritus - 0.21
* (mos$Detritus^2) + 0.04 * (mos$Detritus^3)), size = mos$Egg_Count, log =
T))
## [1] 1415.63
```

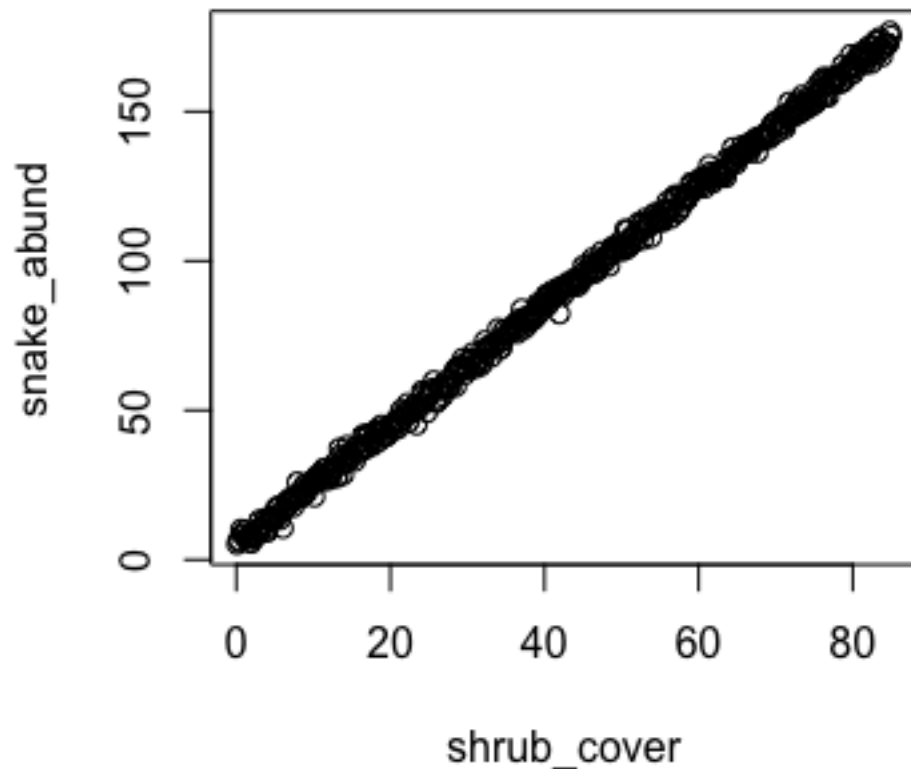
According to dbinom, the likelihood of the data is higher for which model?

The Ricker model has a lower value which means it fits the data better.

#### 4. Power Analysis

a. Linear regression: Percentage of shrub cover predicts snake abundance

```
intercept <- 5
slope <- 2
sample_size <- seq(from = 1, to = 500)
shrub_cover <- seq(from = 0.1, to = 85, length = length(sample_size))
snake_abund <- rnorm(n = sample_size, mean = intercept + slope * shrub_cover,
sd = 2)
plot(snake_abund ~ shrub_cover)
```

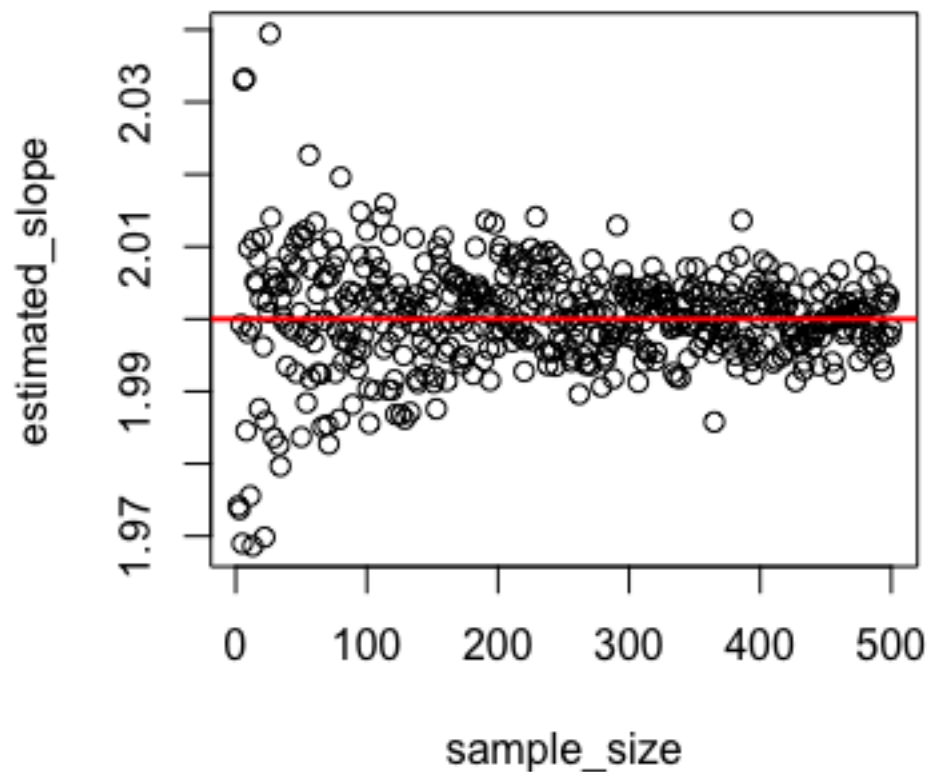




```

sample_size <- seq(from = 1, to = 500)
estimated_slope <- rep(NA, times = length(sample_size))
for(i in 1:length(sample_size)){
  snk_abnd <- rnorm(n = sample_size[i], mean = intercept + slope * seq(from =
0.1, to = 85, length = sample_size[i]), sd = 2)
  mod_snake <- glm(snk_abnd ~ seq(from = 0.1, to = 85, length =
sample_size[i]))
  estimated_slope[i] = coef(mod_snake)[2]
}
plot(estimated_slope ~ sample_size)
abline(h = 2, col = "red", lwd = 2)

```



```
summary(mod_snake)$coefficients[2,4]
```

```
## [1] 0
```

b. Binomial: Percentage of shrub cover predicts the number of recaptured snakes

```

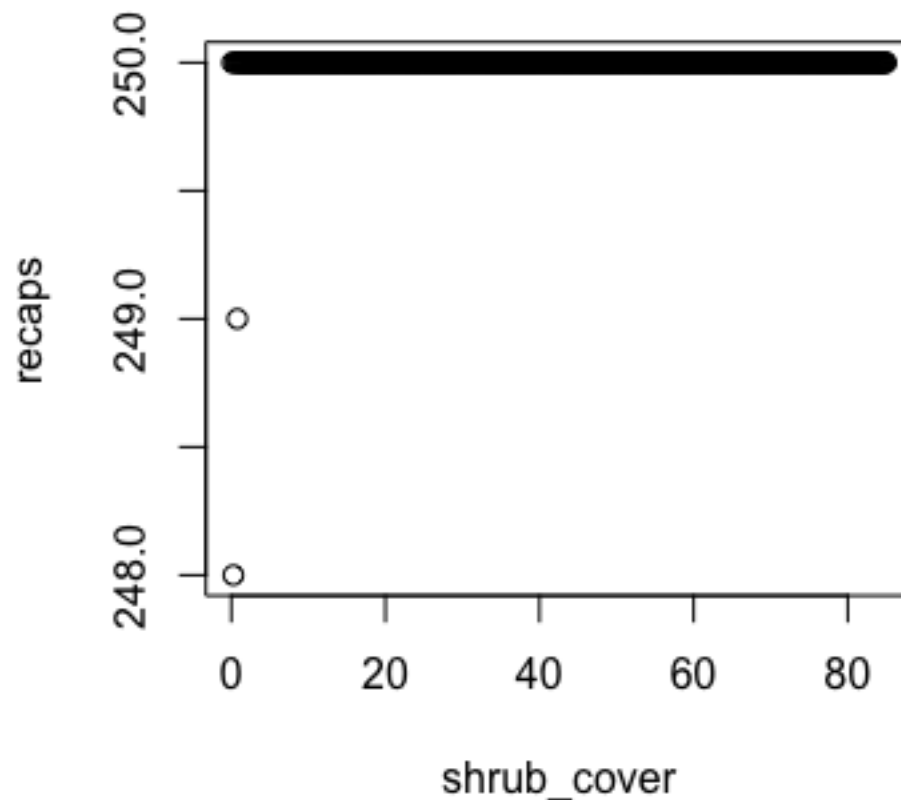
intercept <- 5
slope <- 2
sample_size <- seq(from = 1, to = 500)

```

```

shrub_cover <- seq(from = 0.1, to = 85, length = length(sample_size))
recaps <- rbinom(n = sample_size, prob = plogis(intercept + slope *
shrub_cover), size = 250)
plot(recaps ~ shrub_cover)

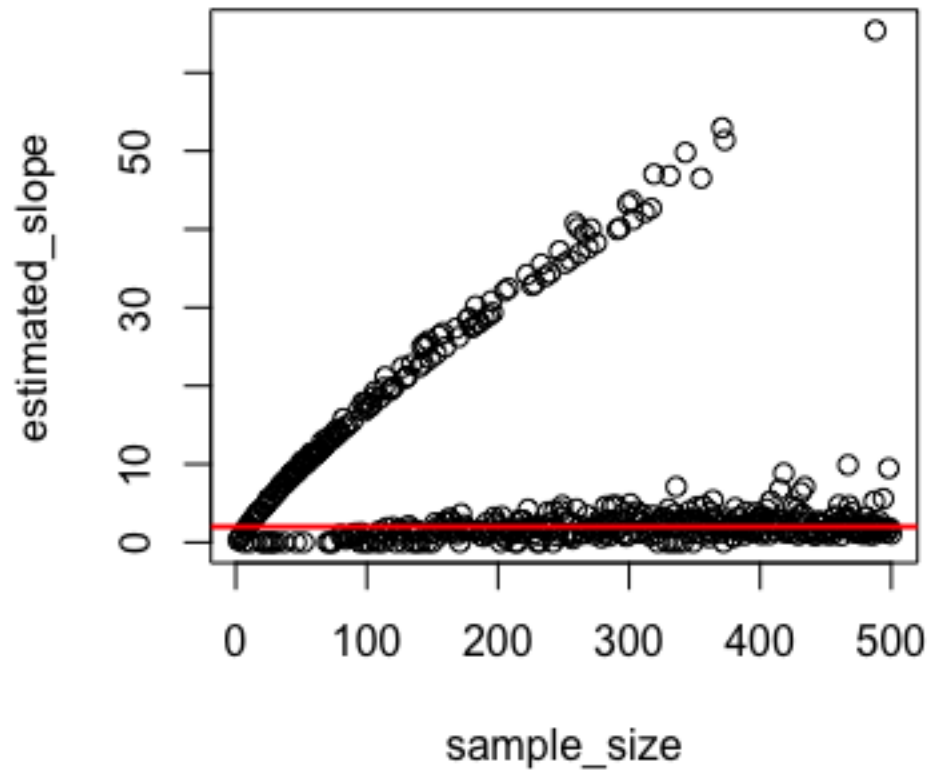
```



```

sample_size <- rep(seq(from = 1, to = 500), times = 1)
estimated_slope <- rep(NA, times = length(sample_size))
for(i in 1:length(sample_size)){
  snk_recap <- rbinom(n = sample_size[i], prob = plogis(intercept + slope *
seq(from = 0.1, to = 85, length = sample_size[i])), size = 250)
  response <- cbind(snk_recap, 250 - snk_recap)
  mod_recap <- glm(response ~ seq(from = 0.1, to = 85, length =
sample_size[i]), family = "binomial")
  estimated_slope[i] = coef(mod_recap)[2]
}
plot(estimated_slope ~ sample_size)
abline(h = 2, col = "red", lwd = 2)

```



```
summary(mod_recap)$coefficients[2,4]
```

```
## [1] 0.317966
```

- a. How many samples do you need to accurately estimate the slope parameter in a binomial vs. linear regression?

For my slope parameter for a linear regression I would need a sample size of at least 234 individual snakes to determine the effect shrub cover has on captures. The estimated slope is getting much closer together as sample size increases. For my binomial model the slope parameters estimate a population size of around 500 recaptured snakes.  $\hat{\theta}^2$

b. How many samples do you need to ensure a  $p\_value < 0.05$  for binomial vs. linear regression?

For binomial model I would need a sample size of around 500 recaptures to ensure a significant p-value. For the linear regression I need a sample size of 234 individuals to have a significant p-value.

- c. In general, why is statistical power generally higher for continuous than discrete response variables?

Because continuous variable allows for less variation between the data. Where as in discrete variables there are gaps between and a lot of variation can occur between these. Less variation means that it is more likely to predict accurately and have more significance.