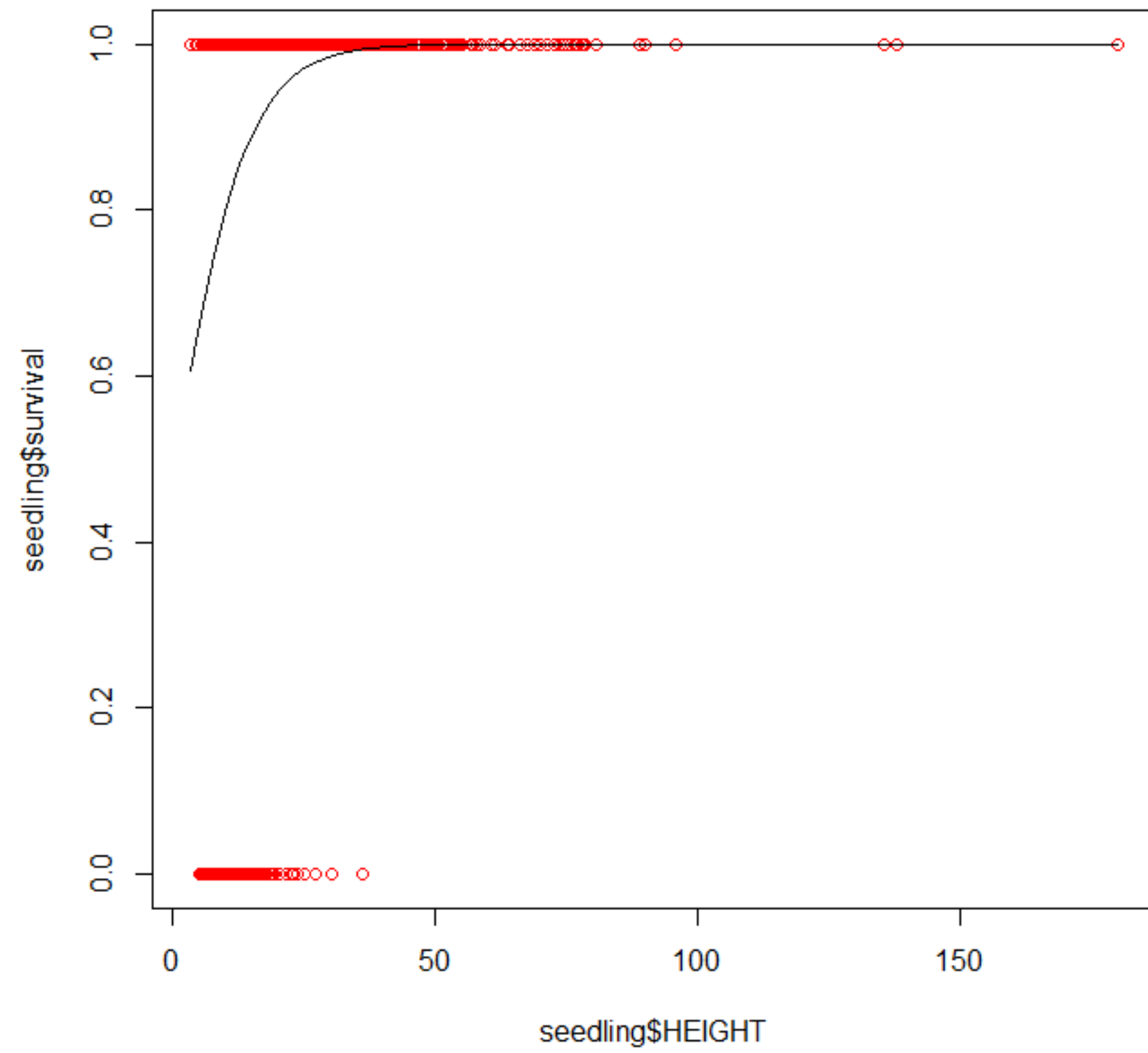


```

> setwd("H:/EEB697/homework2-master/homework2-master")
>
> ##### MSE Function #####
> MSE=function(y_hat,y)
+ {
+   return(mean((y-y_hat)^2))
+ }
> #####
> #* [-----] *#
> #* [----- Q1 -----] *#
> #* [-----] *#
>
> # ##### Seedling Survival vs Height amount #####
>
> seedling <- read.csv("SEEDLING_SURVIVAL.csv") # Reading the raw data
> summary(seedling) # Summary of the data
  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
>
> dev.new() # Creating new window for plot
NULL
> plot(seedling$survival~seedling$HEIGHT,col="red") # Plot raw data (seedling survival vs Height change)
> glm_height<- glm(seedling$survival~seedling$HEIGHT,family = "binomial") # Computing the intercept and slope parameters
> Intercept_height <- coef(glm_height)[1] # Height Intercept
> Slope_height <- coef(glm_height)[2] # Slope of Linear fit
> curve(plogis(Intercept_height+Slope_height*x),add=T) # adding "plogis(a+b*x)" to raw data plot
> plogis(Intercept_height) # The Baseline
(Intercept)
0.4843274
> Slope_height/4 # effect size, by 1 unit increase in height seedling survival increases 3.5%
seedling$HEIGHT
0.03517785
> confint(glm_height) # intercept and slope confident interval
waiting for profiling to be done...
      2.5 %    97.5 %
(Intercept) -0.5791061 0.4268167
seedling$HEIGHT 0.1038803 0.1815477
Warning messages:
1: glm.fit: fitted probabilities numerically 0 or 1 occurred
2: glm.fit: fitted probabilities numerically 0 or 1 occurred
3: glm.fit: fitted probabilities numerically 0 or 1 occurred
4: glm.fit: fitted probabilities numerically 0 or 1 occurred
5: glm.fit: fitted probabilities numerically 0 or 1 occurred
6: glm.fit: fitted probabilities numerically 0 or 1 occurred
7: glm.fit: fitted probabilities numerically 0 or 1 occurred
8: glm.fit: fitted probabilities numerically 0 or 1 occurred
>
> #* [ Note: According to confidence interval, the Intercept crosses the zero ]*#
> #* [ and Binomial does not seem to be the proper distribution for fitting ]*#
> #* [ to our data ]*#
>

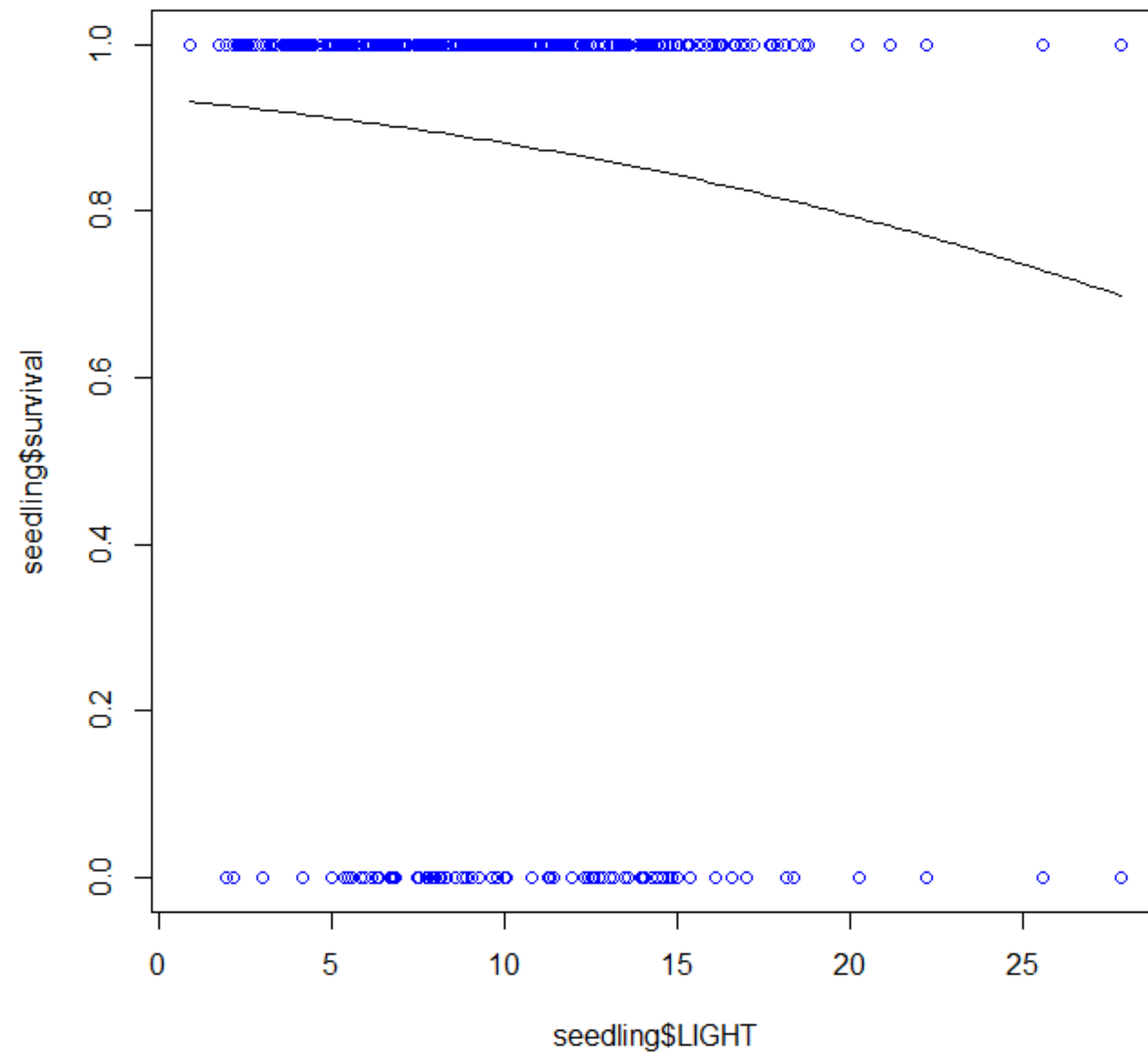
```



```

> # ##### Seedling Survival vs Light amount #####
>
> dev.new() # Creating new window for plot
NULL
> plot(seedling$survival~seedling$LIGHT,col="blue") # Plot raw data (seedling survival vs Light change)
> glm_light<- glm(seedling$survival~seedling$LIGHT,family = "binomial") # Computing the intercept and slope parameters
> Intercept_light <- coef(glm_light)[1] # Light Intercept (a)
> Slope_light <- coef(glm_light)[2] # Slope of Linear fit (b)
> curve(plogis(Intercept_light+Slope_light*x),add=T) # adding "plogis(a+b*x)" to raw data plot
> plogis(Intercept_light) # The Baseline
(Intercept)
0.9347435
> Slope_light/4 # effect size, by 1 unit increase in light seedling survival decreases 1.6%
seedling$LIGHT
-0.01638171
> confint(glm_light) # intercept and slope confident interval
waiting for profiling to be done...
2.5 % 97.5 %
(Intercept) 2.25136434 3.0876309
seedling$LIGHT -0.09841747 -0.0325795
>
> #*[ Note: According to confidence interval Binomial model for Light effect ]*#
> #*[ on seedling survival seems to fit the data in a reasonable way ]*#
>

```



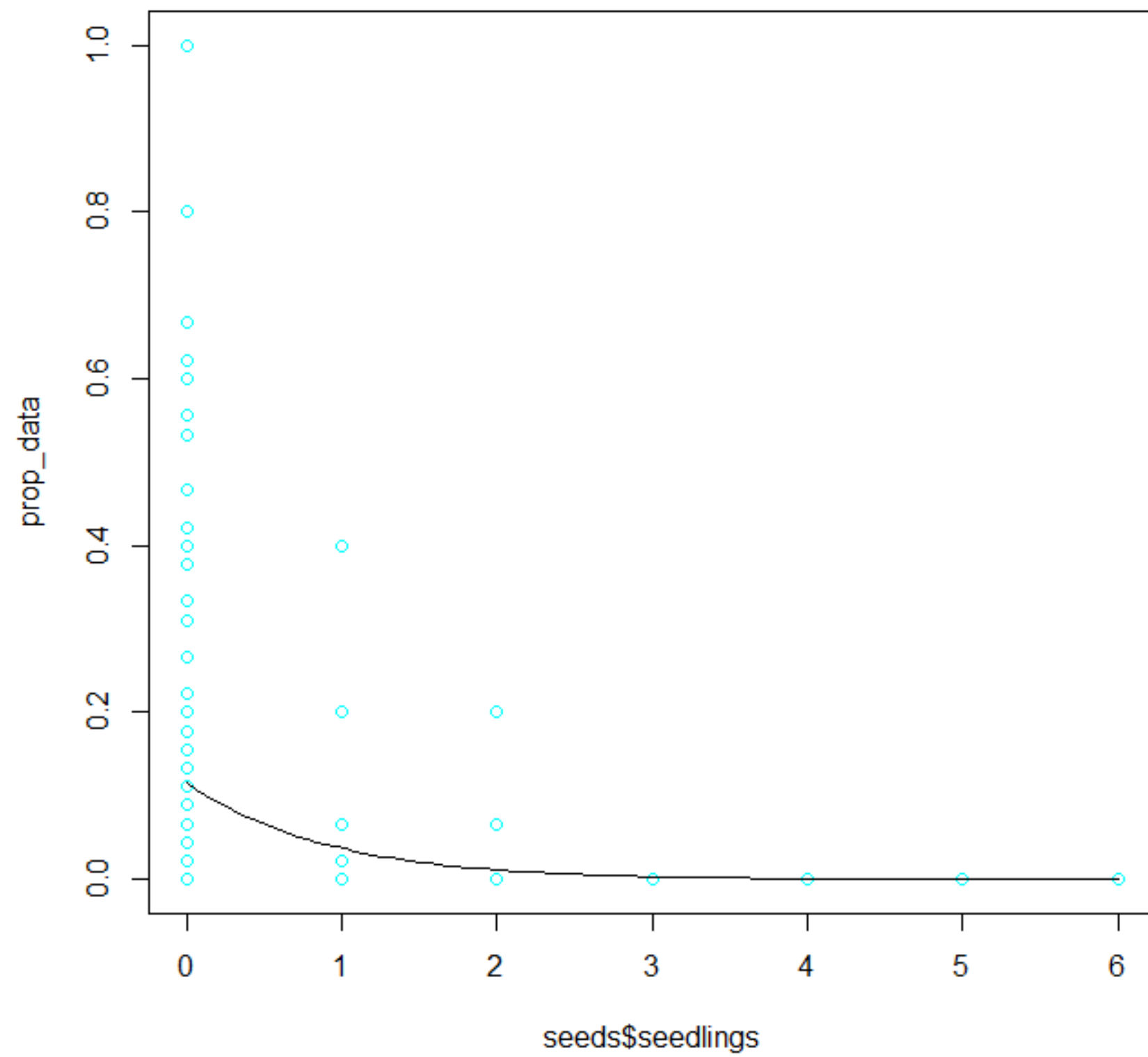
```

> ##[-----]*#
> ##[                Q2                ]*#
> ##[-----]*#
>
> seeds <- read.csv("Seeds.csv")
> summary(seeds)
      Site      Pile      DBH      seedlings      seeds      recruits      grass      light
m1       : 3    m1.15 : 1    Min.       : 0.00    Min.      :0.0000    Min.      : 5.00    Min.      : 0.000    Min.      :0.0000    Min.      : 1.400
m10      : 3    m1.45 : 1    1st Qu.: 0.00    1st Qu.:0.0000    1st Qu.: 5.00    1st Qu.: 0.000    1st Qu.:0.0000    1st Qu.: 3.670
m11      : 3    m1.5  : 1    Median : 0.00    Median :0.0000    Median :15.00    Median : 0.000    Median :0.0000    Median : 5.350
m12      : 3    m10.15: 1    Mean   : 20.27   Mean   :0.3096    Mean   :21.73    Mean   : 2.189    Mean   :0.4306    Mean   : 6.027
m13      : 3    m10.45: 1    3rd Qu.: 31.75   3rd Qu.:0.0000    3rd Qu.:45.00    3rd Qu.: 2.000    3rd Qu.:1.0000    3rd Qu.: 7.750
m14      : 3    m10.5  : 1    Max.    :153.92   Max.    :6.0000    Max.    :45.00    Max.    :28.000    Max.    :1.0000    Max.    :19.810
(Other):263  (Other):275
      NA's      :72

> 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

> prop_data <- seeds$recruits/seeds$seeds
> dev.new()
NULL
> plot(prop_data~seeds$seedlings,col="cyan")
> response <- cbind(seeds$recruits, seeds$seeds-seeds$recruits)
> predictor <- seeds$seedlings
> glm_seedlings <- glm(response~predictor,family="binomial")
> slope_seedlings <- coef(glm_seedlings)[2]
> intercept_seedlings <- coef(glm_seedlings)[1]
> curve(plogis(intercept_seedlings+slope_seedlings*x),add=T)
> plogis(intercept_seedlings)
(Intercept)
0.1155186
> slope_seedlings/4
predictor
-0.3034291
>
>
> ##[ Note: ]*#
> ##[ This result shows that the amount of conspecific seedlings has significant effect ]*#
> ##[ of seedling germination which 1 unit increase in its amount results 30 percent ]*#
> ##[ decrease in germination chance ]*#
> confint(glm_seedlings)
# intercept and slope confident interval
waiting for profiling to be done...
      2.5 %      97.5 %
(Intercept) -2.121808 -1.9511837
predictor    -1.592362 -0.8977661
>

```



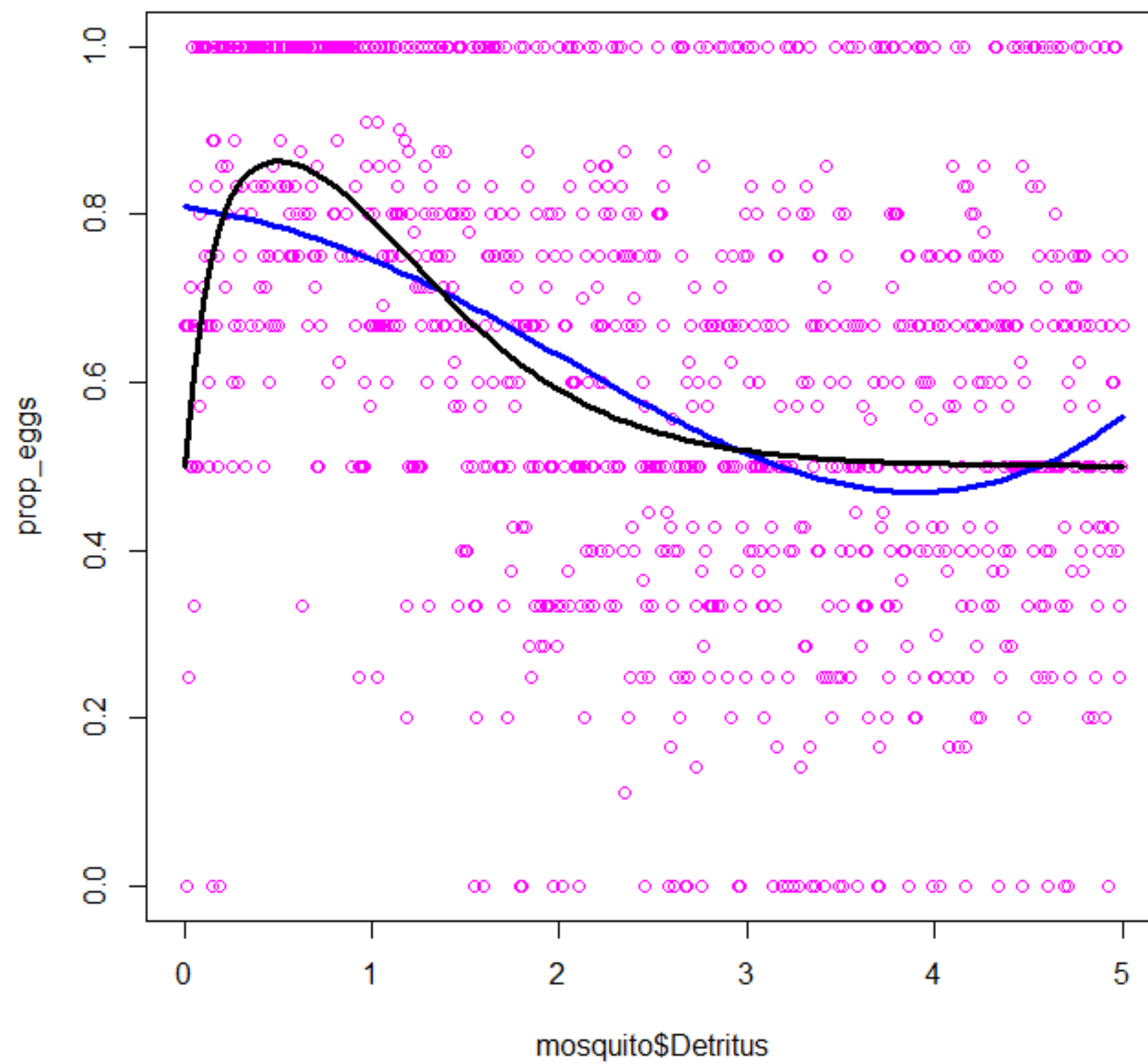
```

> ##[-----]*#
> ##[                Q3                ]*#
> ##[-----]*#
>
> mosquito <- read.csv("mosquito_data.csv")           # Reading the raw data
> summary(mosquito)                                   # Summary of the data
  Emergent_adults  Egg_Count  Detritus
Min.   : 0.000    Min.   : 1.000   Min.   :0.00
1st Qu.: 2.000    1st Qu.: 3.000   1st Qu.:1.25
Median : 3.000    Median : 5.000   Median :2.50
Mean   : 3.063    Mean   : 4.919   Mean   :2.50
3rd Qu.: 4.000    3rd Qu.: 6.000   3rd Qu.:3.75
Max.   :10.000    Max.   :13.000   Max.   :5.00

> head(mosquito)                                     # Header of the data
  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

> prop_eggs <- mosquito$Emergent_adults/mosquito$Egg_Count # Probability (proportional): the amount of adults emerged per egg number
>
> dev.new()                                           # Creating new window for plot
NULL
> plot(prop_eggs~mosquito$Detritus,col="magenta")     # Plot raw data (Probability vs mosquito detritus)
> curve(plogis(1.44-0.19*x-0.21*x^2+0.04*x^3),add=T, col="blue",lwd=3) # Adding "plogis(polynomial)" to raw data plot
> curve(plogis(10*x*exp(-2*x)),add=T,lwd=3)           # Adding "plogis(exponential)" to raw data plot
>
> ##[ Note :                                         ]*#
> ##[ In polynomial function the probability first decreases from 0 to 4 and then ]*#
> ##[ it increases for the detritus values more than ]*#
> ##[ 4. However, in exponential function after reaching maximum probability ]*#
> ##[ between 0 and 1 detritus, it decreases and become fix at 0.5 probability. ]*#
> ##[ exponential distribution shows that by increasing organic ]*#
> ##[ detritus more larva emerge as adult and as this increase continuous ]*#
> ##[ the amount of adult mosquito decreases and after some level organic amounts ]*#
> ##[ in the water does not play any role in adult mosquito population. ]*#
> ##[ In polynomial function, the increase in organic detritus in water first ]*#
> ##[ results in less adult mosquito and then it makes the water environment suitable ]*#
> ##[ for mosquito population increase. ]*#
>
> # Finding the Binomial Likelihood for both polynomial and exponential functions (this code minimizes the negative log-likelihood )
> Likelihood_exp <- -sum(dbinom(x=mosquito$Emergent_adults,size=mosquito$Egg_Count,
+                             prob=plogis(10*mosquito$Detritus*exp(-2*mosquito$Detritus)),log=T))
> print(Likelihood_exp)
[1] 1385.847
> Likelihood_poly <- -sum(dbinom(x=mosquito$Emergent_adults,size=mosquito$Egg_Count,
+                             prob=plogis(1.44-.19*mosquito$Detritus-0.21*mosquito$Detritus^2+0.04*mosquito$Detritus^3),log=T))
> print(Likelihood_poly)
[1] 1415.63
>
> ##[ Note :                                         ]*#
> ##[ As the negative likelihood of the exponential is less than polynomial, ]*#
> ##[ it seems that the exponential function fits better the data ]*#
>

```




```

> #* [-----] *#
> #* [          Q4          ] *#
> #* [-----] *#

```

```

> # ##### Binomial #####

```

```

> intercept=50
> slope=-0.5
> sample_size=500
> score <- seq(from=55,to=145,length=sample_size)
> y.smp <- rbinom(n=sample_size,p=logis(intercept+slope*score),size = 10)
> dev.new()

```

```

NULL

```

```

> hist(y.smp)
> dev.new()

```

```

NULL

```

```

> plot(y.smp~score)

```

```

>
# plot raw data (graduated students vs IQ score)

```

```

> sample_size=rep(seq(from=40,to=90),times=7)
> estimated_slope=rep(NA,times=length(sample_size))
> p_val_bionom=rep(NA,times=length(sample_size))
>
# Make random repetitive samples which is used in the following for Loop
# Create an empty vector for pasting estimated slopes in it
# Create an empty vector for pasting estimated P-values in it

```

```

> for(j in 1:length(sample_size)){
+   y=rbinom(n=sample_size[j],
+           prob=plogis(intercept+slope*
+                       seq(from=55,to=145,
+                           length=sample_size[j])),size=5)
+   response=cbind(y,5-y)
+   m1<-glm(response~seq(from=55,to=145,
+                         length=sample_size[j]),family="binomial")
+   estimated_slope[j]=coef(m1)[2]
+   p_val_bionom[j] <- summary(m1)$coefficients[2,4]
+ }

```

```

# make random binomial distribution data for each sample

```

```

# generating response matrix with success and failure
# estimating the slope and intercept value per each sample

```

```

# Slope estimation
# P-value estimation

```

```

There were 50 or more warnings (use warnings() to see the first 50)

```

```

>
> dev.new()
# Creating new window for plot

```

```

NULL
> plot(estimated_slope~sample_size)
> abline(h=-0.5,col="red",lwd=2)
> Binomial_MSE <- MSE(estimated_slope,slope)
> print(Binomial_MSE)
# Plot estimated slope vs sample size
# Adding the real slope as a red line to the previous plot
# Estimating the result's Mean Square Error (MSE)
# Print the result's MSE

```

```

[1] 1.089615

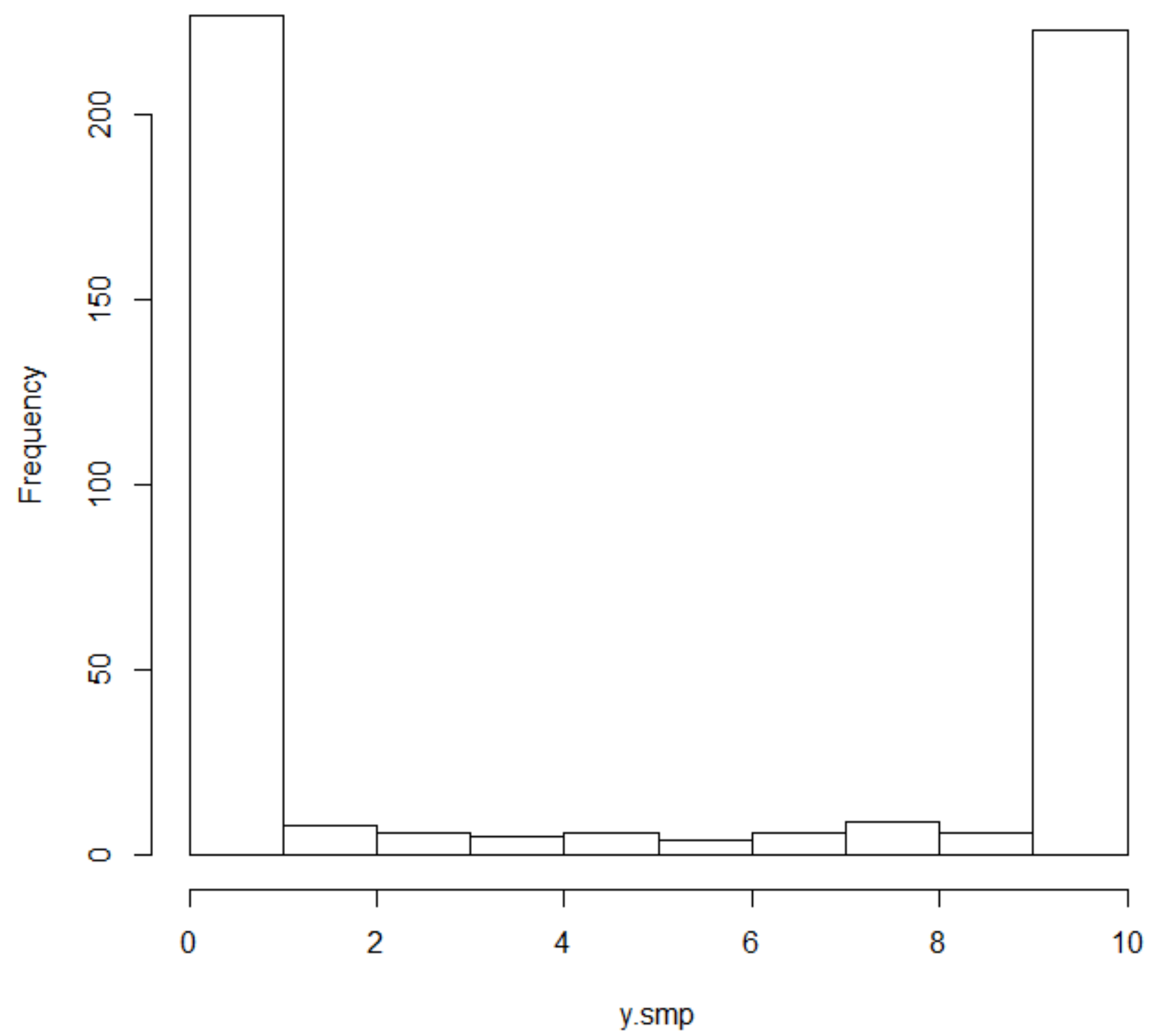
```

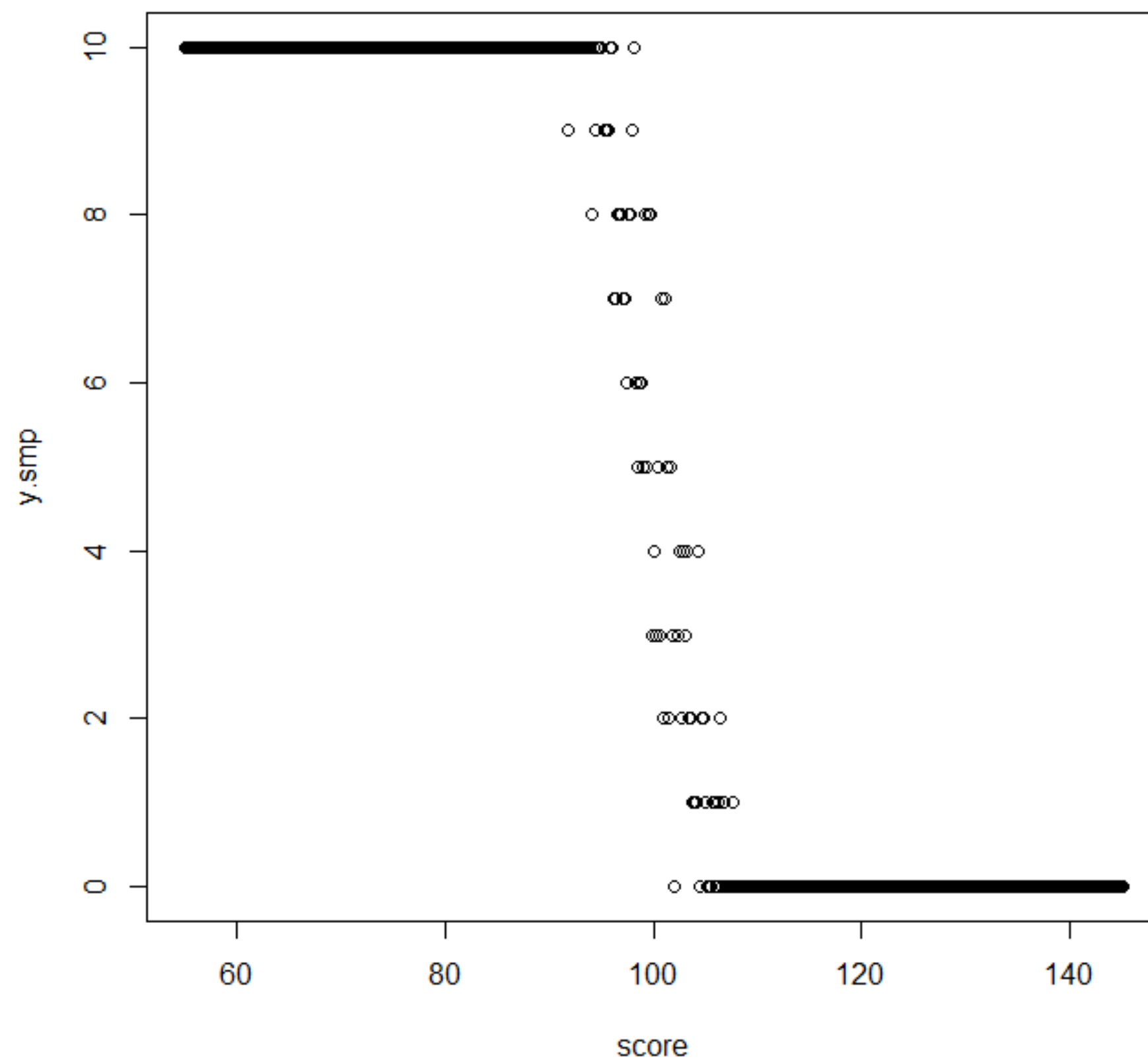
```

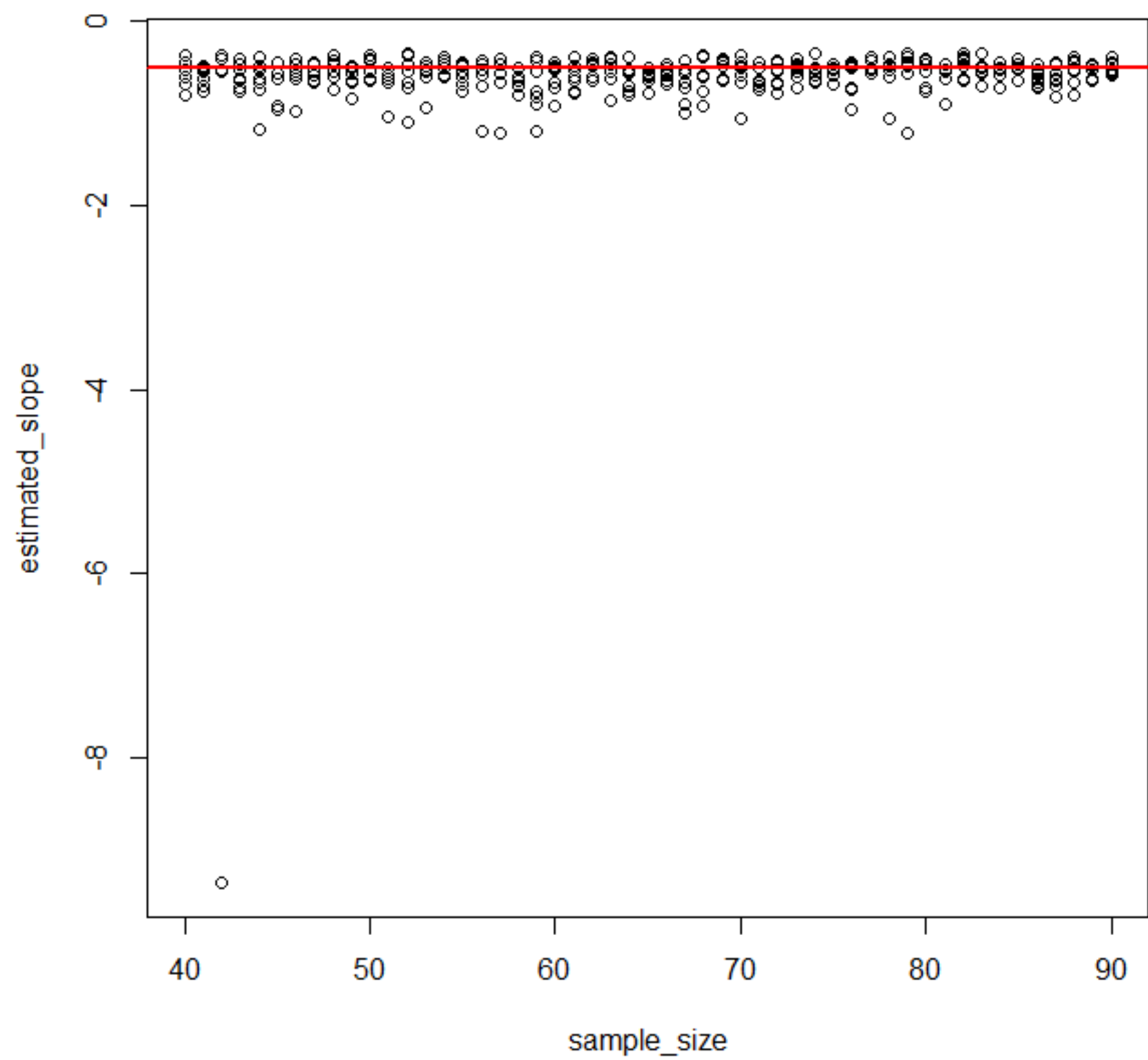
>

```

Histogram of y.smp







Histogram of y_norm

