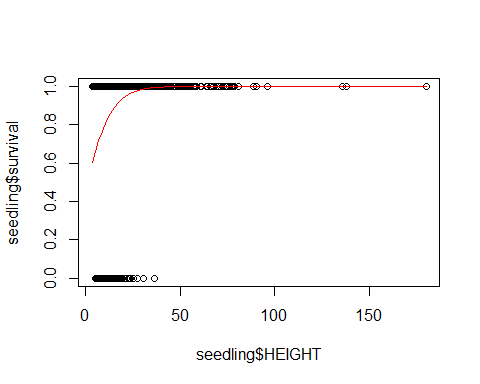
HW2(Jochems)

Louis Jochems

# Question 1   
seedling<-read.csv("SEEDLING\_SURVIVAL.csv")  
head(seedling)

## survival HEIGHT LIGHT  
## 1 1 47.0 2.40  
## 2 1 70.2 14.83  
## 3 1 16.3 9.15  
## 4 1 23.5 8.62  
## 5 1 23.0 4.26  
## 6 1 21.0 3.38

#A) Height on Seedling Survival



m1<-glm(seedling$survival~seedling$HEIGHT,family="binomial")  
summary(m1)

##   
## Call:  
## glm(formula = seedling$survival ~ seedling$HEIGHT, family = "binomial")  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1785 0.1743 0.4765 0.6071 1.0012   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.06271 0.25674 -0.244 0.807   
## seedling$HEIGHT 0.14071 0.01984 7.093 1.31e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1095.35 on 1434 degrees of freedom  
## Residual deviance: 997.64 on 1433 degrees of freedom  
## AIC: 1001.6  
##   
## Number of Fisher Scoring iterations: 7

coef(m1)

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

#a   
plogis(-0.06271111)

## [1] 0.4843274

#b  
0.14071141/4

## [1] 0.03517785

#CI  
#plogis for intercept, real numbers for slope  
plogis(confint(m1))

## Waiting for profiling to be done...

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

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

confint(m1)

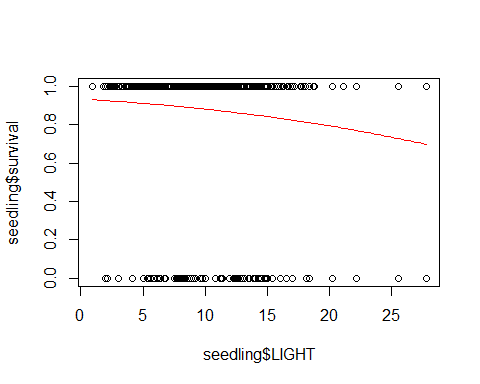
## Waiting for profiling to be done...

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

#b) At maximum effect size, for every unit increase in height there   
#is a 0.133 increase in chance of survival for seedling survival.   
#The coefficient is 0.48, meaning that there is a 48% chance of survival when canopy height is zero.  
  
#c) The confidence interval for the intercept lies between 35 to 60%,   
#indicating that there isuncertainty in the model for baseline survival of seedlings.   
#The CI interval for the slope is tighter, between 2 to 4.5%, indicating   
#that maximum effect of canopy height on seedling survival is significant (does not cross zero) and lies between 2 to 4.5% increase in survival.

#B) Light on Seedling Survival



m2<-glm(seedling$survival~seedling$LIGHT,family="binomial")  
summary(m2)

##   
## Call:  
## glm(formula = seedling$survival ~ seedling$LIGHT, family = "binomial")  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2849 0.4376 0.4908 0.5512 0.8471   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.66195 0.21315 12.488 < 2e-16 \*\*\*  
## seedling$LIGHT -0.06553 0.01677 -3.906 9.37e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1095.4 on 1434 degrees of freedom  
## Residual deviance: 1080.3 on 1433 degrees of freedom  
## AIC: 1084.3  
##   
## Number of Fisher Scoring iterations: 4

coef(m2)

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

#a  
plogis(2.66194692)

## [1] 0.9347435

#b  
-0.06552684/4

## [1] -0.01638171

#CI  
#plogis for intercept, real numbers/4 for slope  
plogis(confint(m2))

## Waiting for profiling to be done...

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

confint(m2)

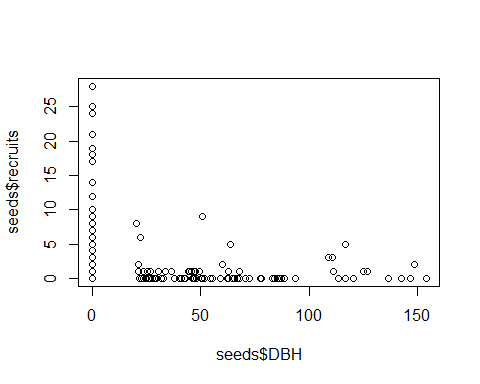
## Waiting for profiling to be done...

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

#b) At maxmimum effect size, we observe a positive slope of 0.12, or a 12%   
#increase in seedling survival for every 1 unit increase of light. For intercept,  
# there is a 93% percent chance of seedling survival at a baseline light of zero.   
  
  
#c) The CI for intercept lies between 90 to 95.6%, giving us high certainty for baseline seedling survival estimate.   
# The CI for slope ranges from -3 and -9%, indicating that light is a significant predictor of seedling survival.   
  
##Answer  
# Based on the CI, we see that height is a stronger predictor of seedling survival than the CI of light. However, the tighter CI (narrower range) for the intercept/slope for light gives us more certainty that the glm parameters are more likely to be correct. Additionally the   
#model paramaters for height fit the data for seedling survival than the model parameters for light.

###question 2###  
seeds<-read.csv("Seeds.csv")

plot(seeds$recruits~seeds$DBH)



#but need to plot data as proportion of seeds suriving out of seeds added  
Survivors<-seeds$recruits/seeds$seeds

#setting up a binomial glm:   
#a) DBH  
seed\_success<-cbind(seeds$recruits,seeds$seeds-seeds$recruits)  
head(seed\_success)

## [,1] [,2]  
## [1,] 2 13  
## [2,] 2 43  
## [3,] 1 4  
## [4,] 0 15  
## [5,] 0 45  
## [6,] 2 3

m1<-glm(seed\_success~seeds$DBH,family="binomial")  
summary(m1)

##   
## Call:  
## glm(formula = seed\_success ~ seeds$DBH, family = "binomial")  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.6595 -1.6123 -0.8911 0.4493 7.4964   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.829832 0.045977 -39.8 <2e-16 \*\*\*  
## seeds$DBH -0.026250 0.002345 -11.2 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1509.1 on 280 degrees of freedom  
## Residual deviance: 1286.0 on 279 degrees of freedom  
## AIC: 1626.1  
##   
## Number of Fisher Scoring iterations: 5

coef(m1)

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

#a  
plogis(-1.82983245)

## [1] 0.1382582

#b  
-0.02625002/4

## [1] -0.006562505

#CI  
#plogis for intercept, real numbers for slope  
plogis(confint(m1))

## Waiting for profiling to be done...

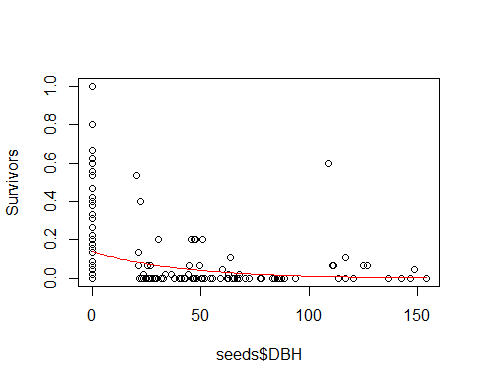
## 2.5 % 97.5 %  
## (Intercept) 0.1277602 0.1492303  
## seeds$DBH 0.4922377 0.4945388

confint(m1)

## Waiting for profiling to be done...

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

#a)  
plot(Survivors~seeds$DBH)  
x<-seeds$DBH  
curve(plogis(-1.82983245+-0.02625002\*x),add=T, col="red")



#b)The intercept estimate for this model is 0.138, or close to a 14% chance of survival for recruits (out of number of seeds added) when DBH of conspecific adults is zero. The estimate for slope is -0.006, indicating that the max effect of conspecific DBH is small but significant, with a decrease of 0.65 % in seed recruitment for every unit increase in conspecific DBH.   
  
#c) The CI for intercept is between 12 and 14% giving us certainty that the estmited baseline survival value is correct. For DBH, the CI is between -0.03 and -0.021, indicating certainty for a small, but significant effect.   
  
# ANSWER:   
# My results do support that my selected predictor variable, conspecifc DBH in plots, has a significant effect on seedling germination given a strong p value, narrow CI intervals for intercept (certainty of parameters) and CI for slope that do not cross zero.

# Question 3   
mosquito<- read.csv("mosquito\_data.csv")  
head(mosquito)

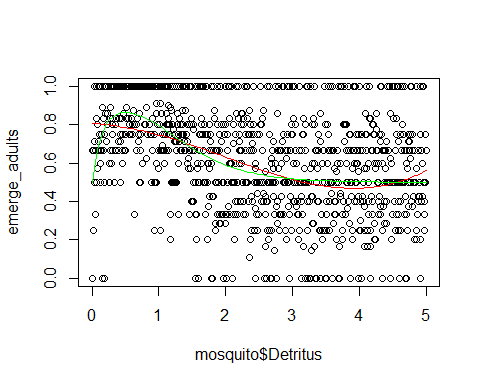
## Emergent\_adults Egg\_Count Detritus  
## 1 2 3 0.00  
## 2 0 2 0.01  
## 3 2 3 0.01  
## 4 4 6 0.02  
## 5 1 4 0.02  
## 6 5 7 0.03

emerge\_adults <- mosquito$Emergent\_adults/mosquito$Egg\_Count  
  
adult\_success<-cbind(mosquito$Emergent\_adults,mosquito$Egg\_Count-mosquito$Emergent\_adults)  
  
m3<-glm(adult\_success~mosquito$Detritus,family="binomial")  
coef(m3)

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

plogis(1.3240425)

## [1] 0.7898535



#c) We see that the polynomial model shows much more variation between emerged adults and org. detritus. The porpotion of emergent adults steadily decreases as amount of org. detritus in pools increases, up until around 4.15 [unit] detritus, where it starts to steadily increase again. This might indicate that detritus may not be a significant predictor of emergent adult success and/or varies under the polynomial model. The Ricker model shows that survival of emergent adults increases steeply from 0 to 1 units of detritus but then steadily decreases and has no effect (levels off around 3 units), perhaps after detritus surpasses some threshold of sufficient abundance.   
  
  
#d) trying to minimize negative log likelihood  
  
-sum(dbinom(mosquito$Emergent\_adults,mosquito$Egg\_Count,   
plogis(10\*mosquito$Detritus\*e^(-2\*mosquito$Detritus)),log=T))

## [1] 1385.847

-sum(dbinom(mosquito$Emergent\_adults,mosquito$Egg\_Count,   
 prob=plogis(1.44-0.19\*x-0.21\*x^2+0.04\*x^3),log=T))

## [1] 1415.63

# ANSWER: The likelihood is higher for the ricker model (1385.85, or minimized negative log likelihood) than for the polynomial model (1415.63). Thus we are more likely to observe the data given the hypothesis under the ricker model.

#question 4   
#stochastic simulation for linear regression   
slope<- -0.1  
intercept<- 350  
  
sample\_size=rep(seq(from=400,to=43), times=10)  
estimated\_slope<-rep(NA,length(sample\_size))  
power\_vector<- rep(NA,3580)  
for(j in 1:length(sample\_size)) {  
 power\_vector=rnorm(sample\_size[j],intercept+slope\*seq(from=500,to=1500,length=sample\_size[j]),25)  
 m1<-glm(power\_vector~seq(from=500,to=1500,length=sample\_size[j]))  
 estimated\_slope[j]=coef(m1)[2]}  
summary(m1)$coefficients[2,4]

## [1] 7.058418e-08

#reps  
358\*10

## [1] 3580

coef(m1)

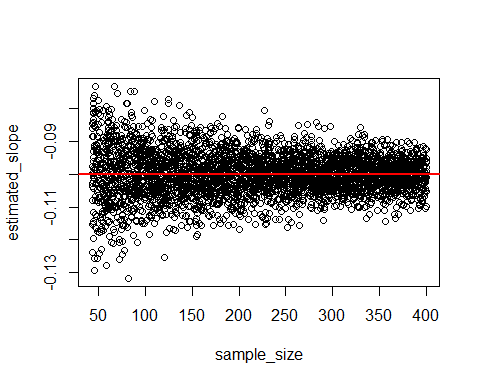
## (Intercept)   
## 332.8311035   
## seq(from = 500, to = 1500, length = sample\_size[j])   
## -0.0843973

confint(m1)

## Waiting for profiling to be done...

## 2.5 %  
## (Intercept) 306.5125919  
## seq(from = 500, to = 1500, length = sample\_size[j]) -0.1096371  
## 97.5 %  
## (Intercept) 359.14961505  
## seq(from = 500, to = 1500, length = sample\_size[j]) -0.05915748

plot(estimated\_slope~sample\_size)  
abline(h=-0.1,col="red",lwd=2)



#binomial regression simulation  
slope<- -0.001  
intercept<- -0.04  
  
sample\_size<- c(3:1000)  
estimated\_slope<-rep(NA,times=length(sample\_size))  
power\_vector<-rep(NA,times=997)  
for(j in 1:length(sample\_size)){  
 power\_vector=rbinom(sample\_size[j],plogis(intercept+slope\*seq(from=1, to=5000,length=sample\_size[j])),size=1)  
 m1<-glm(power\_vector~seq(from=1, to=5000,length=sample\_size[j]),family="binomial")  
 estimated\_slope[j]=coef(m1)[2]}

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

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

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

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

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

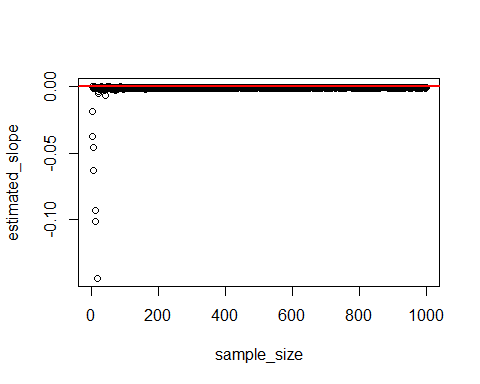
## Warning: glm.fit: algorithm did not converge

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

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

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

plot(estimated\_slope~sample\_size)  
abline(h=0,col="red",lwd=2)



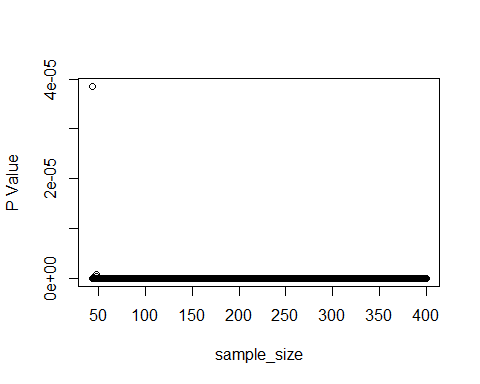
#a) MSE values closer to zero are better  
# for linear regression:   
  
#LR\_MSE  
mean(-0.1-(estimated\_slope))^2

## [1] 0.009695635

#[1] 2.495188e-08  
  
#BR\_MSE  
mean(-0.5-(estimated\_slope))^2

## [1] 0.2484688

#[1] 0.25  
  
#In my linear regression simulation, it looks like I need around 215 samples to ensure accurate estimate parameters. For my binomial regression, it looks like I need between 75 and 100 samples to ensure accurate slope and intercept.   
  
#b)   
#linear regression:   
slope<- -0.1  
intercept<- 350  
sample\_size=rep(seq(from=400,to=43), times=10)  
estimated\_slope<-rep(NA,length(sample\_size))  
power\_vector<- rep(NA,3580)  
for(j in 1:length(sample\_size)) {  
 power\_vector=rnorm(sample\_size[j],intercept+slope\*seq(from=500,to=1500,length=sample\_size[j]),25)  
 m1<-glm(power\_vector~seq(from=500,to=1500,length=sample\_size[j]))  
 estimated\_slope[j]=summary(m1)$coefficients[2,4]}



#binomial regression:   
slope<- -0.001  
intercept<- -0.04  
sample\_size<- c(3:1000)  
estimated\_slope<-rep(NA,times=length(sample\_size))  
power\_vector<-rep(NA,times=length(sample\_size))  
for(j in 1:length(sample\_size)){  
 power\_vector=rbinom(sample\_size[j],plogis(intercept+slope\*seq(from=1, to=5000,length=sample\_size[j])),size=1)  
 m1<-glm(power\_vector~seq(from=1, to=5000,length=sample\_size[j]),family="binomial")  
 estimated\_slope[j]=summary(m1)$coefficients[2,4]}

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

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

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

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

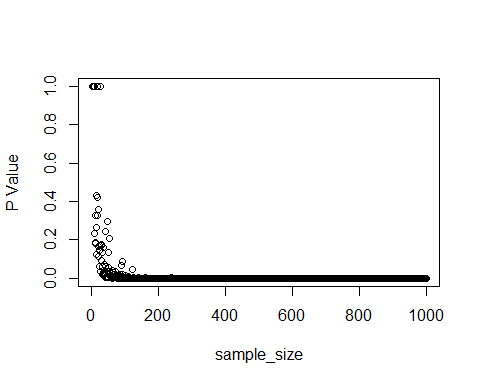
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

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

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

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

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



#My extracted p values are both very small and significant for both binomial and linear regression simulations. However, there are big differences between the estimated vs. actual parameters for binomial model, but much closer model parameters to the actual slope and intercept for the linear regression.   
  
#c) Continuous variables have more information (range of values) to fit within models, generating more statistical power. Discrete response variables have categories that may not fit perfectly within models.