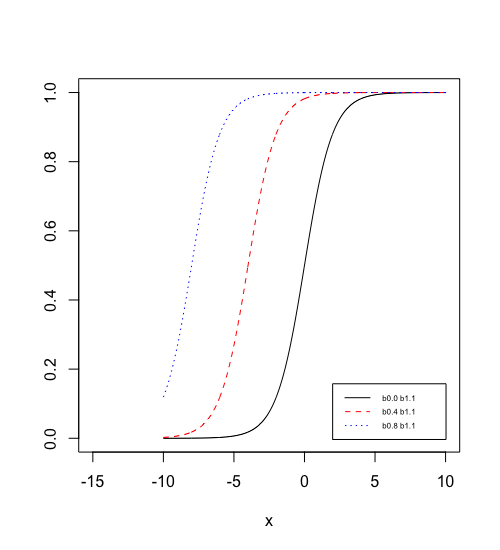
Alyssa Semerdjian

Stat 510

12 Feb 2018

Homework 4

1. B0 determines the upper limit of the logistic curve. The curves will approach but will not pass 1. See code and figure below.



slm=function(b0,b1,x){

fu=1/((1+exp(-(b0+(b1\*x)))))

return(fu)

}

x=seq(-10,10,0.02)

b0=c(0,4,8)

b1=c(1,1,1)

nam1=paste0("b0.",b0[1]," b1.",b1[1])

nam2=paste0("b0.",b0[2]," b1.",b1[2])

nam3=paste0("b0.",b0[3]," b1.",b1[3])

b=slm(b0[1],b1[1],x)

c=slm(b0[2],b1[2],x)

d=slm(b0[3],b1[3],x)

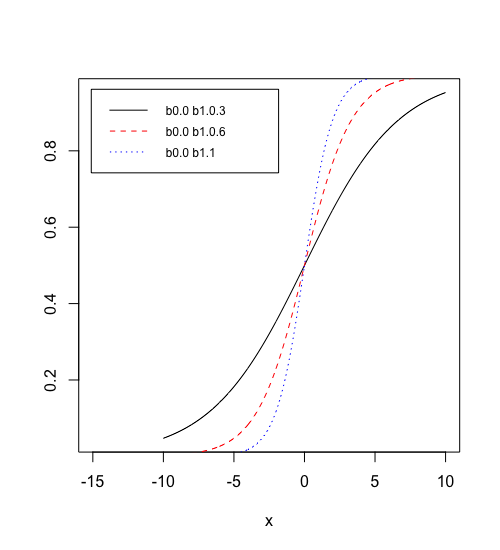
plot(x,b,ylab="",type="l",lty=1,xlim=c(-15,10))

lines(x,c,ylab="",type="l",lty=2,col=2)

lines(x,d,ylab="",type = "l",lty=3,col=4)

legend(locator(1),lty=c(1,2,3),col=c(1,2,4),c(nam1,nam2,nam3),cex=.5)

1. B0 controls the steepness of the curve. A larger value for B0 will result in a line that approaches 1 quicker. See graph below. Code is the same as above with substituted values.



2.

a) The formula for odds ratio is O = p/(1-p). The odds of a married person using marijuana can be expressed as O = 0.1062/(1-0.1062) which is 0.119. The odds of an unmarried person using marijuana = 0.246/(1-0.246) or 0.327. The odds ratio for unmarried vs married marijuana use is calculated using oddsunmarried/oddsmarried = 0.327/0.119 = **2.75**. This means that the odds of marijuana use among the unmarried are 2.75 times the odds of the married. Alternatively, the odds of marijuana use among the married are 0.3638 times the odds of the unmarried.

1. You cannot fit a linear model to a binomial response variable as it will not have a normally distributed data. You should use a logistic model.

c) Call:

glm(formula = mjuser ~ married, family = "binomial", data = drugtest)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.7518 -0.7518 -0.4739 -0.4739 2.1178

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -1.11911 0.03407 -32.85 <2e-16 \*\*\*

married -1.01109 0.05938 -17.03 <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: 8511.1 on 9096 degrees of freedom

Residual deviance: 8197.7 on 9095 degrees of freedom

AIC: 8201.7

Number of Fisher Scoring iterations: 4

The equation for the logistic model is as follows: . There is a significant difference between married and unmarried people’s marijuana use. p-value for this question is <2e-16.

d)The probability of unmarried people’s marijuana use is 0.2462.

> predict(a,list(married=0),type="response")

1

0.246177

e)The probability of married people’s marijuana use is 0.1062.

> predict(a,list(married=1),type="response")

1

0.1061967

f) The odds ratio tells us how much higher the odds of something happening are for one group versus another. In this case the odds ratio is 0.3638, meaning that married people are 0.3638 times as likely to use marijuana as unmarried people. This matches the odds ratio calculated in part a. The odds ratio is different than the probabilities calculated in d and e.

g) -2 \* loglikelihood(model) = 8197.42. This is very close to the residual deviance given by the summary function for the model, which is 8197.7.

h) > 1-pchisq(8511.1-8197.7,1)

[1] 0

We reject the null hypothesis and conclude that the model is valid. There is a statistically significant difference in marijuana use between married and unmarried people.