A study of 482 initial public offering companies (IPOs) was conducted to determine the characteristics of companies that attract venture capital. Here, the response of interest is whether or not the company was financed by venture capital funds. Several potential predictors are: the face value of the company; the number of shares offered; and whether or not the company was a leveraged buyout. The IPO data set is found in file `ipo.csv`. In this example we consider just one predictor, the face value of the company. Use the log of the face value since it is highly skewed. Results are coded as: Y = 1 if company was financed by venture capital funds; Y = 0, otherwise. It is of interest to predict if a company is financed by venture capital funds, based on the estimated face value of the company.

a) Fit a simple logistic regression.

b) Plot the fitted equation

c) Add a lowess smooth curve to the plot.

d) Fit a second order logistic regression model. Which model fits best?

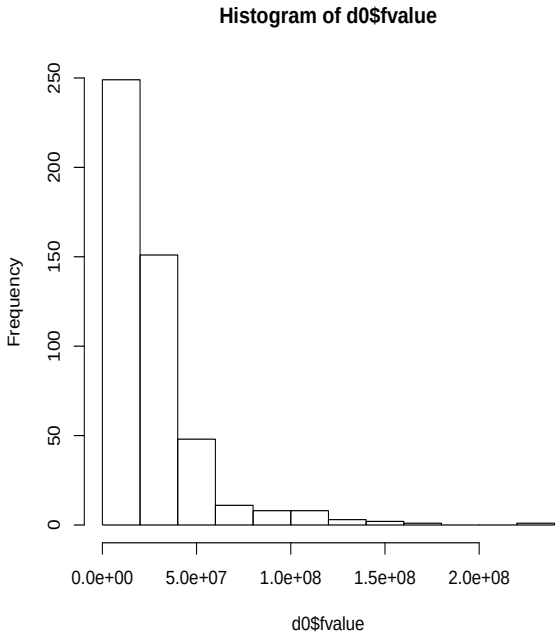
e) Fit a third order logistic regression model. Which model fits best?

f) Use 10-fold cross-validation to find the best prediction model.

```
# ipo6.csv   logistic regression with polynomial terms

rm(list=ls())
setwd("C:/Users/USC Guest/Downloads2")
d0 = read.csv("ipo.csv",header=T)

d1=d0[,c(1,2)]
names(d1)=c("Y","X")
d1$X=log(d1$X)
hist(d1$X)

# simple logistic regression
#================================================================

fit = glm(Y~X,binomial,d1)
summary(fit)
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
#(Intercept)  -7.6722     1.8041  -4.253 2.11e-05 ***
#X             0.4441     0.1075   4.130 3.62e-05 ***

#(Dispersion parameter for binomial family taken to be 1)
#    Null deviance: 661.20  on 481  degrees of freedom
#Residual deviance: 643.13  on 480  degrees of freedom
#AIC: 647.13

summary(d1$X)
# Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#  14.00   16.14   16.79   16.71   17.30   19.27

xx = seq(14,19.27,length=200)
plot(Y~X,d1,pch=19,cex=0.5)


newval=data.frame(X=xx)
yy = predict(fit,newval,type="response")
lines(xx,yy)
grid()

# loess fit
loess = loess(Y~X,d1)
yl = predict(loess, data.frame(X=xx))
lines(xx,yl,lty=2,col="red")


# second order logistic model
#================================================================

fit5 = glm(Y~poly(X,2),binomial,d1)
```

```
summary(fit5)
# Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.4064     0.1088  -3.734 0.000188 ***
poly(X, 2)1  14.2622     2.9099   4.901 9.52e-07 ***
poly(X, 2)2 -21.1088     3.4401  -6.136 8.46e-10 ***

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 661.20  on 481  degrees of freedom
Residual deviance: 588.27  on 479  degrees of freedom
AIC: 594.27

# plot
xx = seq(14,19.27,length=200)
plot(Y~X,d1,pch=19,cex=0.5)
newval=data.frame(X=xx,X2=xx^2)
yy = predict(fit5,newval,type="response")
lines(xx,yy)
grid()

# loess fit
lines(xx,yl,lty=2,col="red")
grid()

# third order logistic model
#=================================================================

fit6 = glm(Y~poly(X,3),binomial,d1)
summary(fit6)
# Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.3728     0.1082  -3.445 0.000571 ***
poly(X, 3)1  10.8744     3.5743   3.042 0.002347 **
poly(X, 3)2 -19.0091     3.5268  -5.390 7.05e-08 ***
poly(X, 3)3  -5.9173     4.2976  -1.377 0.168549

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 661.20  on 481  degrees of freedom
Residual deviance: 586.52  on 478  degrees of freedom
AIC: 594.52

xx = seq(14,19.27,length=200)
plot(Y~X,d1,pch=19,cex=0.5)
newval=data.frame(X=xx,X2=xx^2,X3=xx^3)
yy = predict(fit6,newval,type="response")
lines(xx,yy)
grid()

lines(xx,yl,lty=2,col="red")
grid()
```

```
# k-Fold Cross-Validation
#============================================================
# Leave 10-out models

library(boot)
set.seed(17)
mspe=rep(0,5)   # initialize vector

for (i in 1:5)
{
 models=glm(Y~poly(X,i),binomial,d1)
 mspe[i]=cv.glm(d1,models,K=10)$delta[1]
}

mspe
# 0.2399499 0.2142489 0.2157637 0.2170965 0.2153225
```



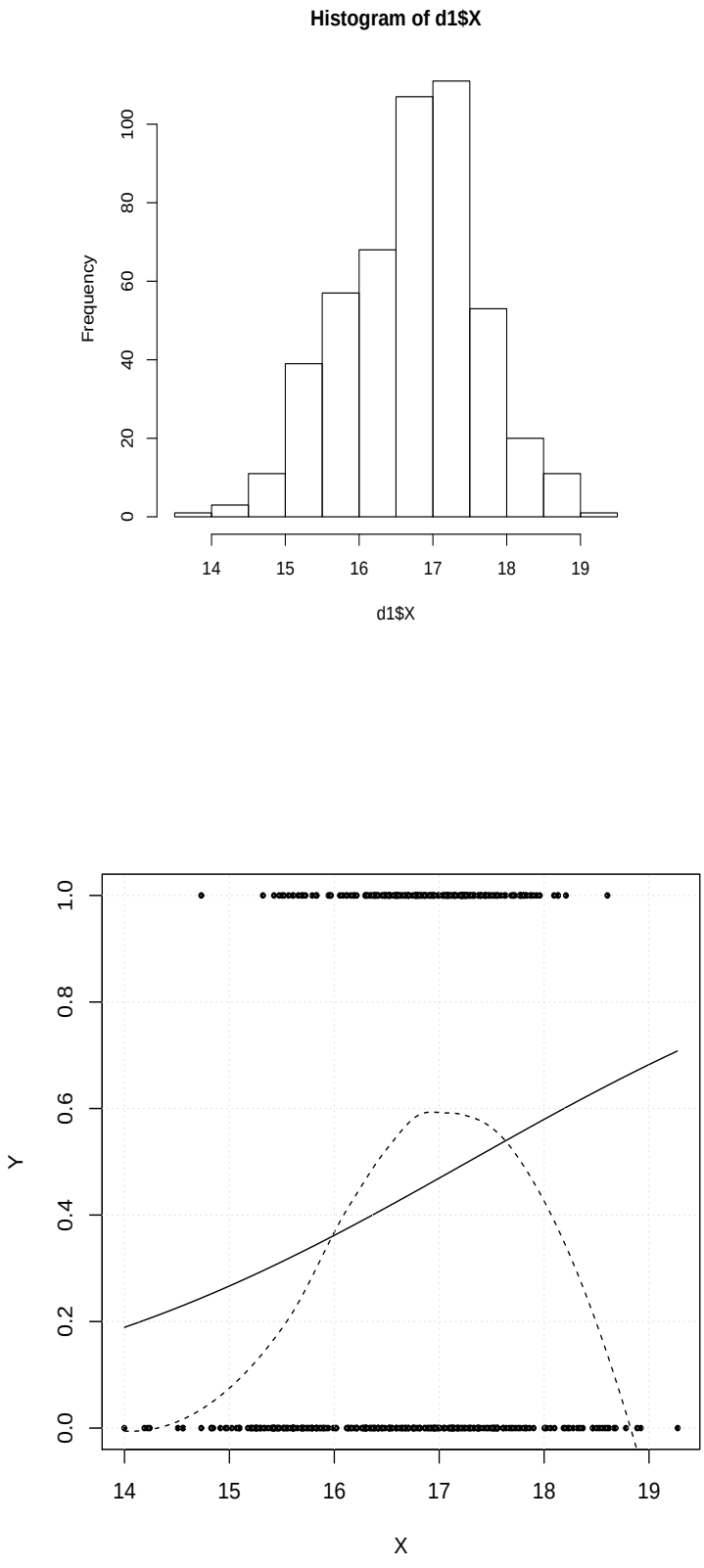


Figure 1: Simple logistic regression

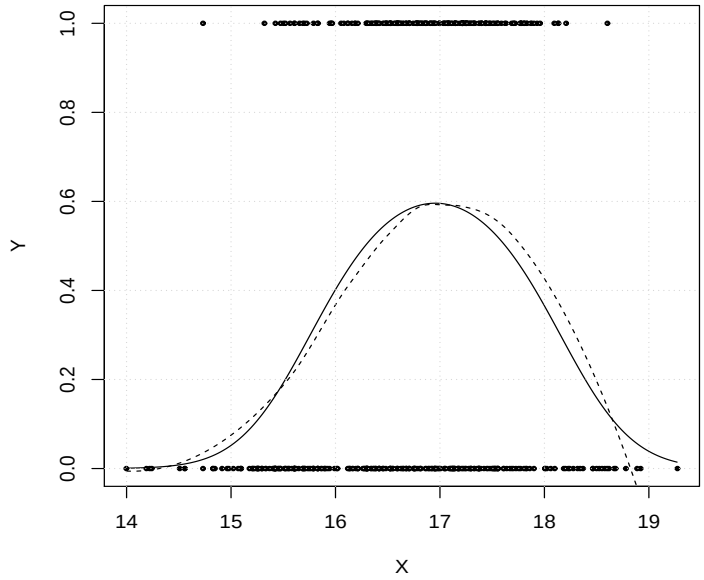


Figure 2: Second order polynomial logistic regression

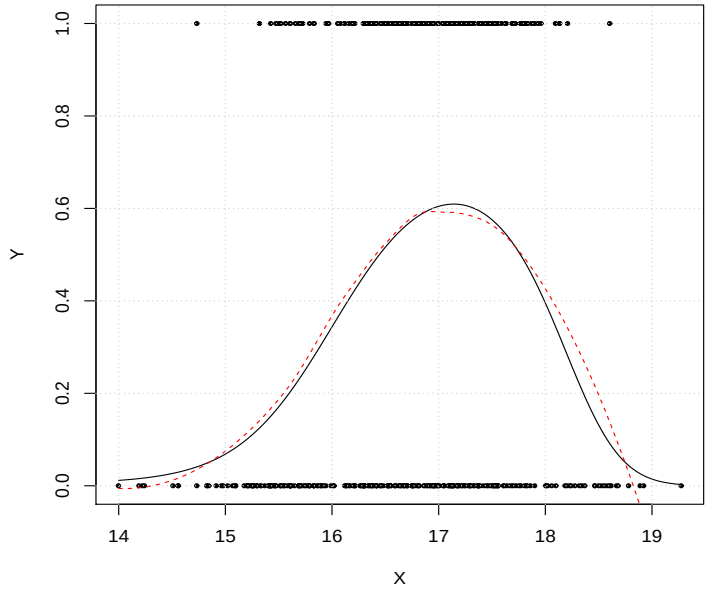


Figure 3: Third order polynomial logistic regression