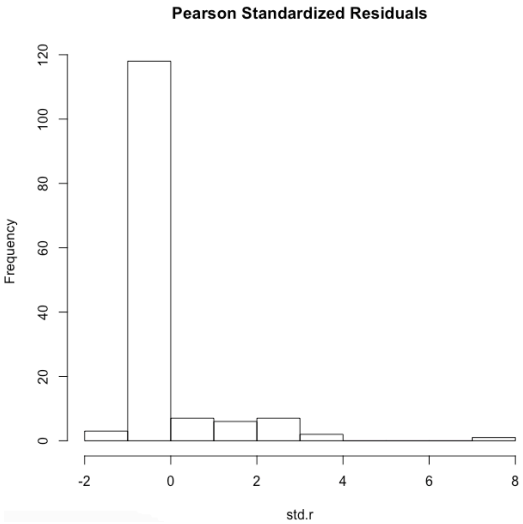
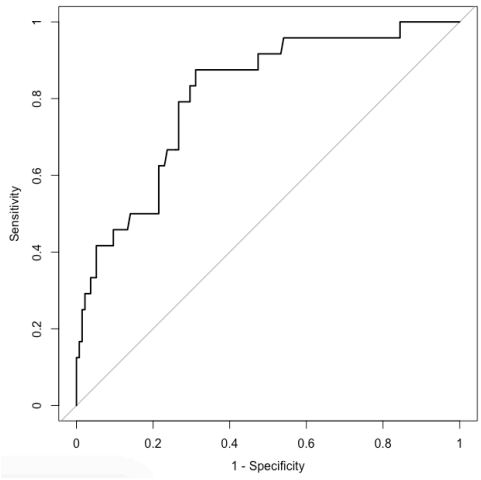
Kathleen Zhen

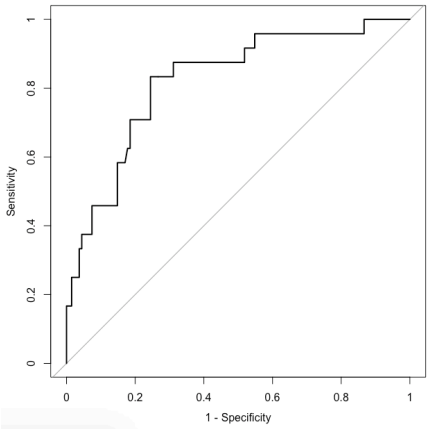
9b



10a.

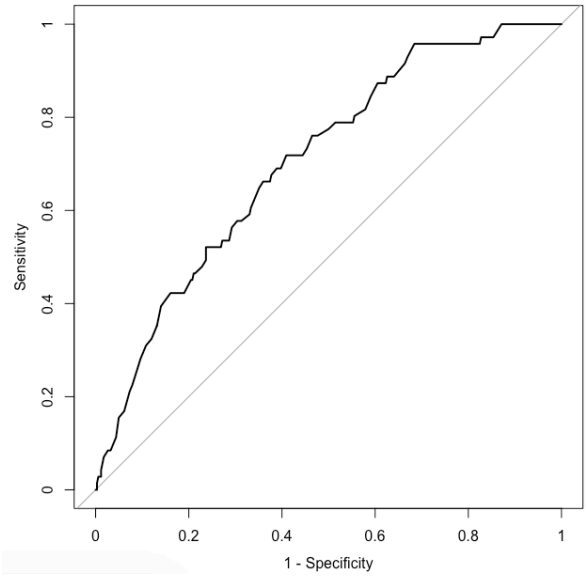


10c.

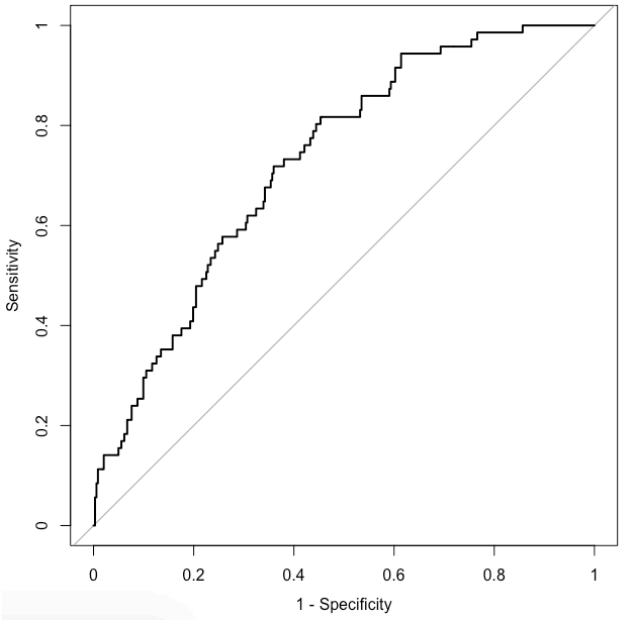


Kathleen Zhen

12c.



12d.



Kathleen Zhen

APPENDIX

```
#1
internet <- read.csv("~/Desktop/internet.csv")

big.logit = glm(Newbie ~ Age + Gender + score, data = internet, family = binomial)
big.logit
summary(big.logit)
exp(-0.06075)
exp(1.1087)
exp(1.02337)

#2
alpha = 0.01
confint(big.logit)
predict(big.logit, newdata = data.frame(Gender = "Female",  Age = 28,  score = 4),type =
"response")

good.stuff = dx(big.logit)
pear.r = good.stuff$Pr #Pearsons Residuals
deviance.r = good.stuff$dr #Deviance Residuals
std.r = good.stuff$sPr #Standardized residuals (Pearson)
df.beta = good.stuff$dBhat
good.stuff[df.beta > 0,]

#3
work <- read.csv("~/Desktop/work.csv")
cat.logit.model = glm(obese~ gender + age + factor(work$marriage),family = binomial, data =
work)
cat.logit.model

estimates =  summary(cat.logit.model)$coefficients[,1] # A vector of only the estimates
SE =  summary(cat.logit.model)$coefficients[,2] #A vector of only the Wald SE's
alpha = 0.01
z.a.2 = qnorm(1-alpha/2)
upper.bounds = estimates +z.a.2*SE
lower.bounds = estimates -z.a.2*SE
Wald.CI = cbind(lower.bounds,upper.bounds)
Wald.CI


big.logit = glm(obese ~ gender, data = work, family = binomial)
big.logit
```

Kathleen Zhen

```
exp(0.6162)

#5
a = glm(obese~ gender + age+ min + factor(marriage ),family = binomial, data = work)
predict(a, newdata = data.frame(gender = 0,  age = 28,  min = 400 , as.factor(1)),type =
"response")
predict(big.logit, newdata = data.frame(gender = 0,  age = 28,  min = 400),type = "response")
summary(a)

#5c
#include min or not for 5b? adn 5a
b = glm(obese~ gender + min,family = binomial, data = work)
predict(b, newdata = data.frame(gender = 0,  min = 400),type = "response")

summary(b)

c = glm(obese ~ min, family = binomial, data = work)
d = glm(obese ~ gender, family = binomial, data = work)

L0 = logLik(d)
L1 = logLik(b)
LR.test = as.numeric(-2*(L0 - L1))
LR.pval = pchisq(LR.test, df = 1,lower.tail = F )
LR.test
LR.pval


#8
flu <- read.csv("~/Desktop/flu.csv")
full.model = glm(shot ~. , data = flu,family = binomial(link=logit))

empty.model = glm(shot~ 1 ,data = flu,family = binomial(link=logit))
best.FB.AIC = step(empty.model,scope = list(lower = empty.model, upper = full.model),direction
= "both", criterion = "AIC", trace = FALSE)
best.FB.AIC$formula
best.FB.BIC = step(empty.model,scope = list(lower = empty.model, upper = full.model),direction
= "both", k = log(nrow(flu)), trace = FALSE)

b = glm(shot~ aware + age,family = binomial(link = logit), data = flu)
predict(b, newdata = data.frame(aware = 84,  age = 54),type = "response")


#9
best.model = best.FB.BIC
```

Kathleen Zhen

```
library(ResourceSelection)
HL.test = hoslem.test(best.model$y, best.model$fitted.values,g = 8)
HL.test

library(LogisticDx)
good.stuff = dx(best.model)
pear.r = good.stuff$Pr #Pearsons Residuals
deviance.r = good.stuff$dr #Deviance Residuals
std.r = good.stuff$sPr #Standardized residuals (Pearson)
df.beta = good.stuff$dBhat #DF Beta for removing each observation
change.pearson = good.stuff$dChisq #Change in pearson X^2 for each observation
change.LR = good.stuff$dDev #Change in LR-test G^2 for each observation

hist(std.r, main = "Pearson Standardized Residuals")
cutoff.pearson = 3
std.r[std.r > cutoff.pearson] #Shows the values
good.stuff[std.r > cutoff.pearson,1:3]
x = sort(df.beta, decreasing = TRUE)
good.stuff[x,]

cutoff.beta = 0.16
df.beta[df.beta > cutoff.beta] #Shows the values
good.stuff[df.beta > cutoff.beta,]

#10
library(pROC)
the.roc = roc(best.model$y, best.model$fitted.values,auc = TRUE, ci = TRUE,plot=TRUE,
legacy.axes = TRUE)
auc(the.roc)
ci(the.roc)

the.roc = roc(full.model$y, full.model$fitted.values,auc = TRUE, ci = TRUE,plot=TRUE,
legacy.axes = TRUE)
auc(the.roc)
ci(the.roc)

#11
largework <- read.csv("~/Desktop/largework.csv")
library(bestglm)

full.model = glm(y~ . ,data = largework,family = binomial(link=logit))
empty.model = glm(y~ 1 ,data = largework,family = binomial(link=logit))
#Forward selection
```

Kathleen Zhen

```
best.forward.AIC = step(empty.model,scope = list(lower = empty.model, upper =
full.model),direction = "forward", criterion = "AIC", trace = FALSE)
best.backward.AIC = step(full.model,scope = list(lower = empty.model, upper =
full.model),direction = "backward", criterion = "AIC", trace = FALSE)
best.FB.AIC = step(empty.model,scope = list(lower = empty.model, upper = full.model),direction
= "both", criterion = "AIC", trace = FALSE)
best.BF.AIC = step(full.model,scope = list(lower = empty.model, upper = full.model),direction =
"both", criterion = "AIC", trace = FALSE)

best.subset.AIC = bestglm(Xy = largework, family = binomial(link=logit),IC = "AIC",method =
"exhaustive")
best.subset.AIC
best.subset.BIC = bestglm(Xy = largework, family = binomial(link=logit),IC = "BIC",method =
"exhaustive")
best.subset.BIC


best.model = best.subset.AIC$BestModel
the.roc = roc(best.model$y, best.model$fitted.values,auc = TRUE, ci = TRUE,plot=TRUE,
legacy.axes = TRUE)
auc(the.roc)
ci(the.roc)

best.model = best.subset.BIC$BestModel
the.roc = roc(best.model$y, best.model$fitted.values,auc = TRUE, ci = TRUE,plot=TRUE,
legacy.axes = TRUE)
auc(the.roc)
ci(the.roc)
```