

Math 536 Exam 2

Lennox Garay

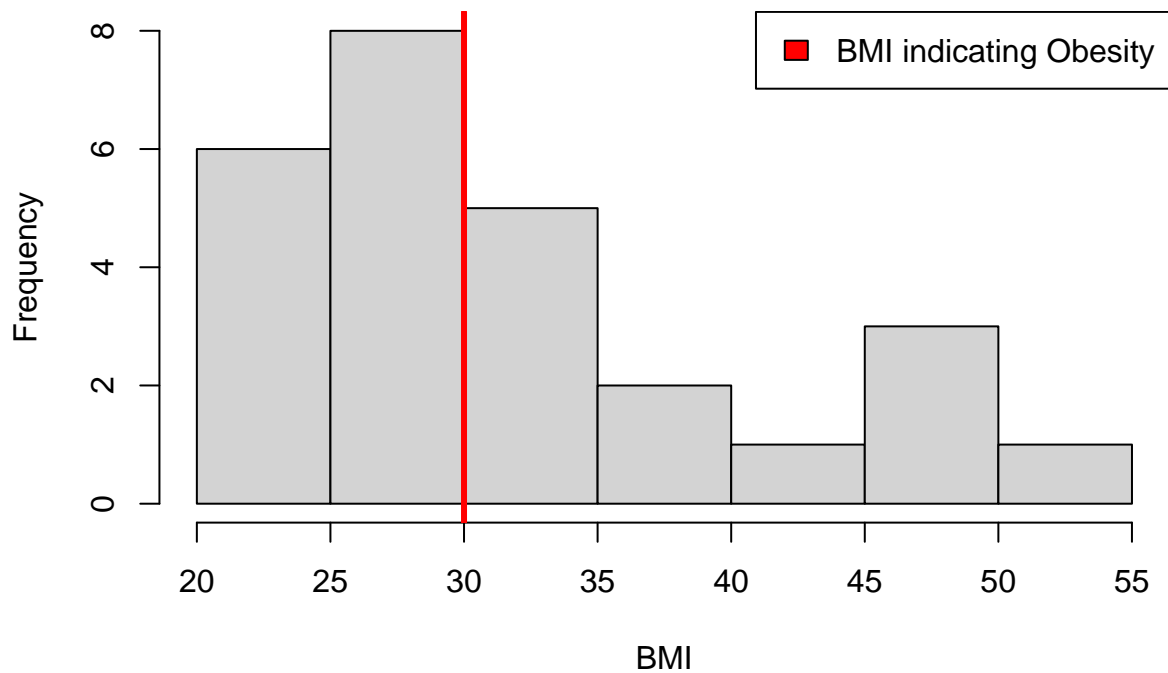
2023-05-01

The Problem:

We have been tasked by research clinicians to investigate possible risk factors of anastomotic leaking, which is a postoperative complication following colorectal surgery.

Specifically, researchers want to investigate the relationship between BMI and increased risk of anastomotic leaking. Below is just an exploration of the data and various ways to illustrate the relationship between BMI and anastomotic leaking.

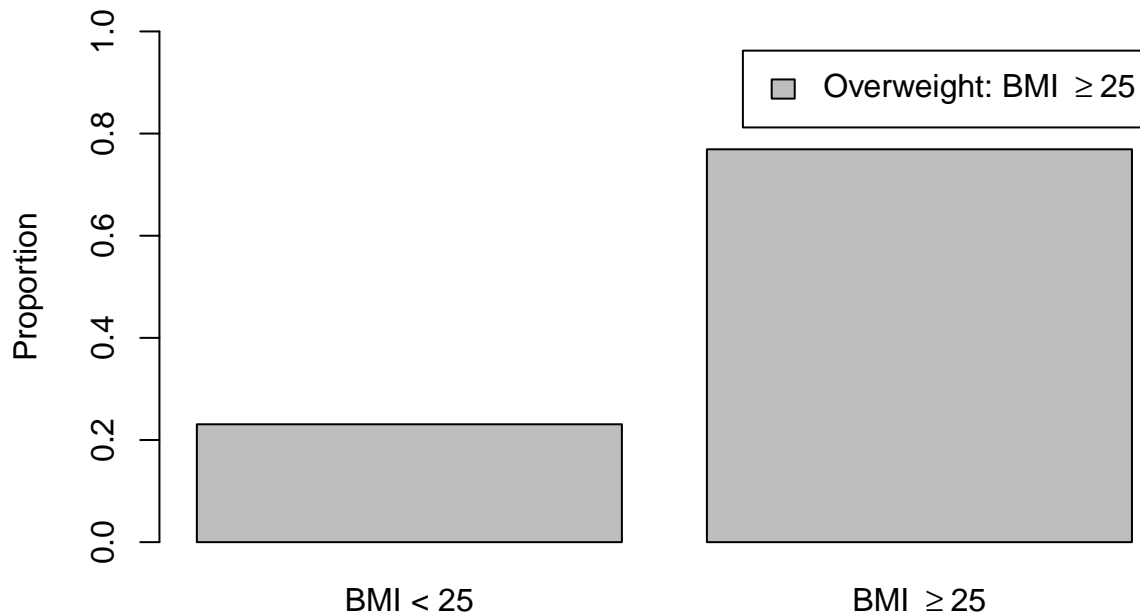
Distribution of BMI in patients with Anastomotic Leak



```
## [1] "BMI under 25 leak proportion:" "0.230769230769231"
```

```
## [1] "BMI at or over 25 leak proportion:" "0.769230769230769"
```

Proportion of BMI to Anastomotic Leak



```
## [1] "Biserial correlation of BMI greather than 25 and Anasttomotic Leak:"  
## [2] "-0.0874505437304073"
```

Investigating the relationship between BMI and oods of anastomotic leaking

53% of the leaks were in people under 30 BMI, which is considered Obese. Obesity itself does not seem to affect leak risk from this measure alone.

When looking at overweight BMI ($BMI > 25$), we see a drastic difference in the proportions. Perhaps looking just obesity is a bit myopic. Considering those with $BMI > 25$, which is still “unhealthy” can pose a great risk for anastomotic leakage.

Still, this is not a complete or succinct explanation of the relationship between BMI and anastomotic leaking. To get the bigger picture of the BMI to leak risk, we look to a generalized linear model of binomial family.

Using response variable (Anastomotic leak), and predictor variables (BMI, Race, CAD/PAD, Albumin, and Operative Length), we get a model coefficient value of 0.05118, which we use to obtain log odds.

Ultimately, we can expect that a 1-unit increase in BMI escalates the risk of anastomotic leaking by 5.25%. 95% of the time the risk of leaking increases between 0.0003% and 10.8% per 1-unit increase in BMI. Considering the ease of increasing BMI can be, it is a pretty significant risk for leaking.

Choosing predictors

Using the AIC, the model best variable selection has reduced the general model to a model containing only the following variables:

Gender, BMI, Age, Tobacco, DM, Albumin, and Operative Length.

These variables were found to be the most important predictors for influencing the anastomotic leaking odds. This AIC variable selection model will not serve us during the case studies. We will construct a new model that contains the predictor values associated with each case study.

The purpose of variable selection was to determine the variables that impact anastomotic leaking the most. The other variables can still provide useful insight for the case study models.

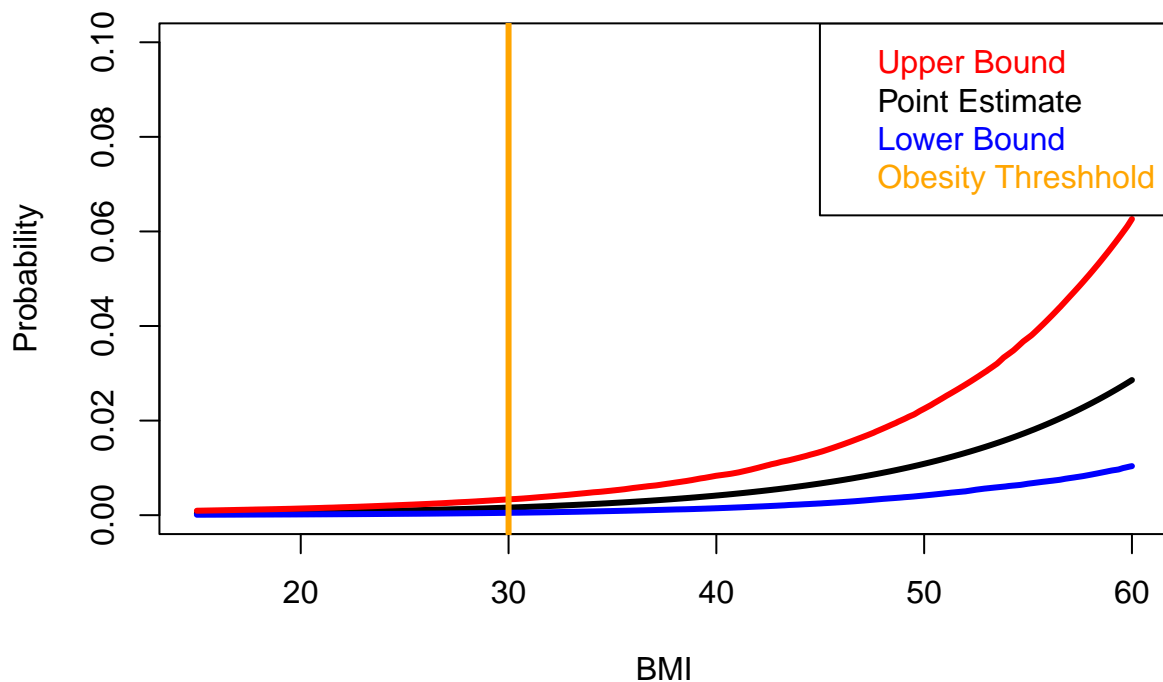
A discussion of the efficiency of the model and potential limitations are discussed at the end of this report under Appendix 1.1:Residuals.

Case study of 2 Patients with Differing Risk Factors

We are presented with 2 patients. The first being ‘Arizona Robbins,’ a 35 year old white female. Arizona does not use tobacco, does not have diabetes, does not have Coronary/Peripheral Artery Disease, does not have cancer, an albumin level of 4.2 and a 90 minute operative length.

This case study is interested in the effect of BMI on the risk of anastomotic leaking. Given Arizona’s profile, we can construct a logistic regression model that demonstrates the relationship between BMI and anastomotic leaking.

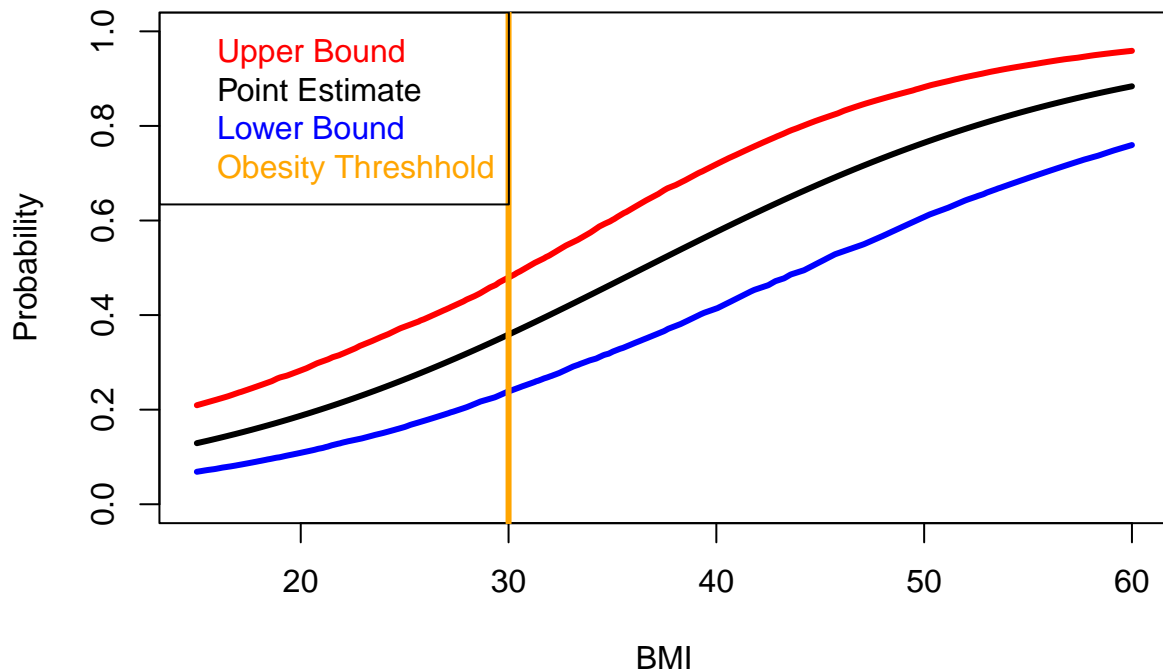
Predicting Probability of Anastomotic Leak using BMI (Arizona)



The graph above shows the point estimate of leaking given a range of BMI values. It shows that Arizona overall is at an extremely low risk for leaking. Even if her BMI was 60, her probability of leaking is still less than 8% given her current health condition.

On the other hand, we have the second person in the case study, Richard Webber. He is a 62 year-old African American Male who uses tobacco and has diabetes. He had an albumin level of 2.8 following a 210 minute operation.

Predicting Probability of Anastomotic Leak using BMI (Richard)



Looking at Richard's graph, we see that his health condition puts him at a great risk for anastomotic leaking. Overall, this would be of significant concern to the doctors overseeing Richard's case.

BMI along with other health statuses greatly increase the odds of getting an anastomotic leak following colorectal surgery.

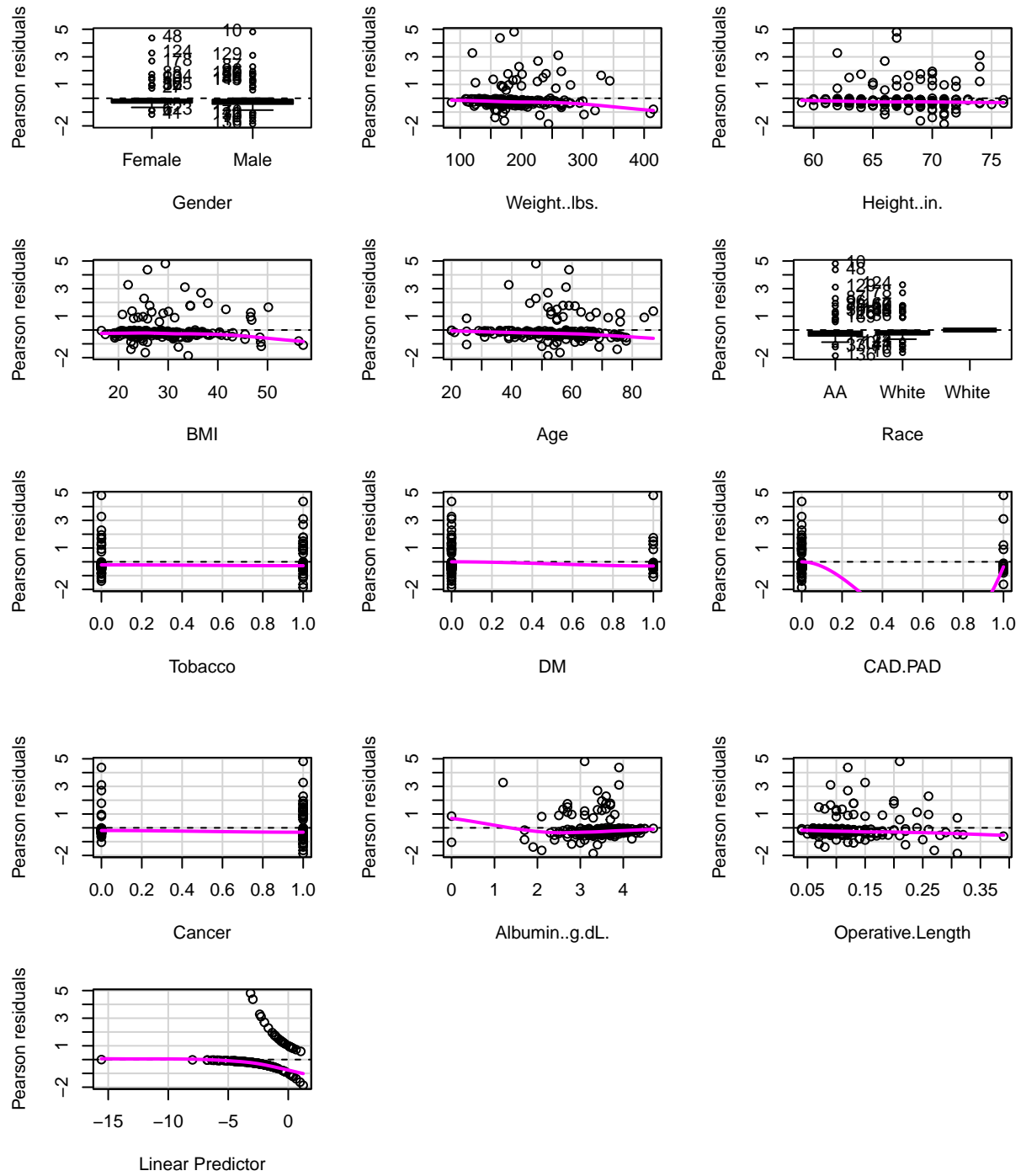
These two cases present polar conditions, which reflect polar probabilities in developing a leak. It's clear that a combination of variables such as tobacco, cancer, diabetes, and operative length (to name a few) can significantly impact whether or not someone is at risk for postoperative complications. Taking a detailed medical history is necessary for evaluating the risk of patients following the surgery.

In detail, we saw that our model performed best when the variables in the model were:

Gender, BMI, Age, Tobacco, DM, Albumin, and Operative Length.

Doctors should keep an eye on patients that have a high BMI, are older in age, use tobacco, have diabetes, have low albumin, and high operative length. These are the direction for the variables associated with an increased risk of anastomotic leaking.

Appendix 1.1: Residuals



```
##              Test stat Pr(>|Test stat|)
## Gender
## Weight..lbs.      3.2615      0.07092 .
## Height..in.       0.1270      0.72160
## BMI               2.6405      0.10417
## Age               0.0064      0.93641
## Race
## Tobacco           0.0000      1.00000
## DM                0.0000      1.00000
## CAD.PAD           0.0000      1.00000
## Cancer            0.0000      1.00000
## Albumin..g.dL.    0.0182      0.89260
## Operative.Length  4.8008      0.02845 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The linear predictor plot shows us that the model tends to over estimates on the higher end of the data set, which is probably fine. It would be better to overestimate a patients odds of leaking than underestimating it, as underestimating could cause doctors to ignore patients that are actually at higher risk. The absence of a follow-up on an at-risk patient could be potentially life threatening.

Appendix 1.2: Coding

```
data = read.csv('colon2017.csv')
data$Race[data$Race == 'W'] <- 'White' ## Race data cleaning
data$Race[data$Race == 'white'] <- 'White'

obese.bmi = filter(data, BMI >= 30) ## table of only obese patients
leak = filter(data, Anastomotic.Leak == 1) ## patients who had a leak
View(leak)
hist(leak$BMI, xlab='BMI', main='Distribution of BMI in patients with Anastomotic Leak'); abline(v=30,
legend("topright", c("BMI indicating Obesity"), fill='red')

h.leak = filter(leak, BMI < 30)$BMI ## People of BMI < 30 who had a leak.
ob.leak = filter(leak, BMI >= 30)$BMI

#### OVERWEIGHT ANALYSIS

norm.leak = filter(leak, BMI < 25)$BMI ## normal bmi patients who had a leak
ow.leak = filter(leak, BMI >= 25)$BMI ## Overweight patients, who had a leak

print(c('BMI under 25 leak proportion:',length(norm.leak)/nrow(leak)))

print(c('BMI at or over 25 leak proportion:',length(ow.leak)/nrow(leak)))

barplot(c(0.2307692,0.7692308),
        ylim=c(0,1),
        names.arg=c('BMI < 25', expression("BMI ">="25")),
        main = 'Proportion of BMI to Anastomotic Leak',
        ylab = 'Proportion',
        legend.text = expression("Overweight: BMI ">="25"))

#### Overweight Correlation
```

```

ow.bmi = filter(data, BMI >= 25)

print(c('Biserial correlation of BMI greather than 25 and Anasttomotic Leak:', biserial.cor(ow.bmi$BMI,

bmi.lm = glm(data$Anastamotic.Leak ~ data$BMI + data$Race + data$CAD.PAD + data$Albumin..g.dL. + data$

summary(bmi.lm)

exp(5.118e-02) ## 1 - exp(5.386) is the 1-unit percent increase

exp(5.118e-02-1.96*2.631e-02);exp(5.118e-02+1.96*2.631e-02)

model = glm(Anastamotic.Leak ~ Gender + Weight..lbs. + Height..in. + BMI + Age + Race + Tobacco + DM +
            family = 'binomial',
            data = data)

summary(model)
step(model)
model.new = glm(Anastamotic.Leak ~ Gender + BMI + Age + Tobacco + DM + Albumin..g.dL. + Operative.Length

summary(model.new)

coef(model.new)

##### ARIZONA MODEL #####
data = read.csv('colon2017.csv')
data$Race[data$Race == 'W'] <- 'White' ## Race data cleaning
data$Race[data$Race == 'white'] <- 'White'
new.data=data[-c(1:4,6,7,10,14,16,17,20)] ## Data containing trimmed model from variable selection

arizona = data.frame(Gender='Female',
                     BMI= seq(15, 60, length.out=1000),
                     Age=rep(35, 1000),
                     Tobacco=rep(0,1000),
                     DM = rep(0,1000),
                     CAD.PAD = rep(0, 1000),
                     Albumin..g.dL.= rep(4.2,1000),
                     Operative.Length=rep(0.05025,1000),
                     Anastamotic.Leak = rep(1,1000))

BS.b1 = rep(0, 1000)
BS.pstar = matrix(0, nrow=1000, ncol=1000)
for(j in 1:1000){
  orig.model = glm(Anastamotic.Leak ~ Gender +
                  BMI +
                  Age +
                  Tobacco +

```

```

        DM +
        CAD.PAD +
        Albumin..g.dL.+
        Operative.Length, data=new.data)

n = 1000
row.index = sample(1:179,1000,replace=T)

BS.x = new.data[row.index,1:8]

log.odds.bs = predict(orig.model, newdata=BS.x)
prob.bs = exp(log.odds.bs)/(1+exp(log.odds.bs))

BS.y = new.data[row.index,9]
newmodel = glm(BS.y ~ Gender + BMI + Age + Tobacco + DM + CAD.PAD + Albumin..g.dL. + Operative.Length,
               data=new.data, family=binomial)

BS.b1[j] = coef(newmodel)[3]
BS.pstar[,j] = predict(newmodel, newdata=arizona[,1:8], type='response')
}

## BS.pstar[,1] == n x m, n = 15 < BMI < 60 / 1000)

lowbound = sort(BS.pstar[,1])[25]
highbound = sort(BS.pstar[,1])[975]

lowbound = rep(0,1000)
highbound = rep(0,1000)
median = rep(0,1000)
mean = rep(0, 1000)
for(i in 1:1000){
  lowbound[i] = sort(BS.pstar[i,])[25]
  highbound[i] = sort(BS.pstar[i,])[975]
  median[i] = sort(BS.pstar[i,])[500]
  mean[i] = mean(BS.pstar[i,])
}

### ARIZONA PLOT
plot(x = seq(15, 60, length.out=1000), y = mean , xlab='BMI', ylab='Probability', ylim = c(0,0.1), main='ARIZONA PLOT')
lines(x= seq(15, 60, length.out=1000), y=lowbound, col='blue', type='l', lwd=3)
lines(x=seq(15, 60, length.out=1000), y=highbound, col='red', type='l', lwd=3)
abline(v=30, col='orange', lwd=3)
legend('topright', legend=c('Upper Bound', 'Point Estimate', 'Lower Bound', 'Obesity Threshold'), text.col=c('red', 'black', 'black', 'black'))

### RICHARD MODEL ###
newdata2 = new.data[-6]

```



```

webber = data.frame(Gender='Male',
                    BMI= seq(15, 60, length.out=1000),
                    Age=rep(62, 1000),
                    Tobacco=rep(1,1000),
                    DM = rep(1,1000),
                    Albumin..g.dL.= rep(2.8,1000),
                    Operative.Length=rep(0.1458333333,1000),
                    Anastamotic.Leak = rep(1,1000))

BS.b1 = rep(0, 1000)
BS.pstar = matrix(0, nrow=1000, ncol=1000)
for(j in 1:1000){
  orig.model = glm(Anastamotic.Leak ~ Gender +
                  BMI +
                  Age +
                  Tobacco +
                  DM +
                  Albumin..g.dL.+
                  Operative.Length, data=newdata2)

  n = 1000
  row.index = sample(1:179,1000,replace=T)

  BS.x = newdata2[row.index,1:7]

  BS.y = newdata2[row.index,8]
  newmodel = glm(BS.y ~ Gender + BMI + Age + Tobacco + DM + Albumin..g.dL. + Operative.Length, family=
  BS.pstar[,j] = predict(newmodel, newdata=webber[,1:7], type='response')
}

lowbound = rep(0,1000)
highbound = rep(0,1000)
median = rep(0,1000)
mean = rep(0, 1000)
for(i in 1:1000){
  lowbound[i] = sort(BS.pstar[i,])[25]
  highbound[i] = sort(BS.pstar[i,])[975]
  median[i] = sort(BS.pstar[i,])[500]
  mean[i] = mean(BS.pstar[i,])
}

### RICHARD PLOT

plot(x = seq(15, 60, length.out=1000), y = mean , xlab='BMI', ylab='Probability', ylim = c(0,1), main='')
lines(x= seq(15, 60, length.out=1000), y=lowbound, col='blue', type='l', lwd=3)
lines(x=seq(15, 60, length.out=1000), y=highbound, col='red', type='l', lwd=3)
abline(v=30, col='orange', lwd=3)
legend('topleft', legend=c('Upper Bound', 'Point Estimate', 'Lower Bound', 'Obesity Threshold'), text.

```

RESIDUALS

```
model = glm(Anastamotic.Leak ~ Gender + Weight..lbs. + Height..in. + BMI + Age + Race + Tobacco + DM +  
            family = 'binomial',  
            data = data)
```

```
residualPlots(model) ## Requires cars package.
```