

Grey's Report

Using method used to analyze the data was a Generalized Linear Model so that all requested measurements may be met. Through the development of the model we realized there were two variables that were not statistically important, thus Cancer and CAD.PAD was dropped in forming the model. This increased our measurement of fit for the model, how well our results are matching the data, so we knew it was safe to drop.

After creating the model we gathered the estimates, standard error, z-value, significance, percent change, along with an upper and lower bound for the percent change to show which variables may be most impactful. The following table shows is for your reference:

Variable:	Estimate:	STD Error:	Z-Value	Pr(> z)
Intercept	-6.921137	2.09214	-3.308154	0.000939
Gender	0.885888	0.551056	1.607618	0.107919
BMI	0.075467	0.032745	2.304711	0.021183
Age	0.078657	0.025078	3.136427	0.00171
Race	-0.540977	0.55694	-0.97134	0.331379
Tobacco	0.504598	0.563257	0.895858	0.370329
DM	-0.879259	0.664124	-1.323938	0.185524
Albumin..g.dL.	-1.637559	0.475041	-3.447196	0.000566
Operative.Length	0.221255	0.158438	1.396479	0.16257

Variable:	%Change	Lower Bound	Upper Bound
Intercept	-99.90132932	-99.99836575	-94.04258799
Gender	142.5137762	-89.61744443	5564.59107722
BMI	7.83877393	-98.82248749	9776.07440336
Age	8.18328559	-99.76859149	50475.59508561
Race	-41.78210977	290.73066197	-91.32568014
Tobacco	65.63190769	-71.38626487	858.76783382
DM	-58.49095326	456.03427284	-96.90126842
Albumin..g.dL.	-80.55459337	16615.89295326	-99.97737938
Operative.Length	24.7642146	-91.92051856	1826.622315

This shows that these variables are all the important metrics in determining the risk of an individual having an anastomotic leak. For every unit increase in BMI we see an increased risk of having an anastomotic leak by %3.2 to %17.1 .

Appendix: Code

```
library(car)
library(boot)
library(stringr)

setwd("C:\\Users\\User\\Desktop\\School\\Math_536\\HW\\HW3")
Data=read.csv("colon2017.csv")

head(Data)

drops=c("Diagnosis.ICD9.Code","CPT.Code","Procedure","Incision.Start","Incision.Close")
Data2=Data[!(names(Data) %in% drops)]
head(Data2)

Data2$Race=as.character(Data2$Race)
fixes=c("White","white","White ","W")

Data2[Data2$Race %in% fixes,][,7]="W"
unique(Data2$Race)

as.character(temp[1,7])==as.character(temp[2,7])
temp[1,7]==temp[2,7]
drops=c("Anastamotic.Leak")

x=Data2[!(names(Data2) %in% drops)]
y=Data2["Anastamotic.Leak"]

Data2$Operative.Length=Data2$Operative.Length*24
```

```
model =  
glm(Anastamotic.Leak~Gender+BMI+Age+Race+Tobacco+DM+CAD.PAD+Cancer+Albumin..g.dL.+Operati  
ve.Length,data=Data2,family="binomial")
```

```
summary(model)
```

```
#Potential drops: Cancer,CAD.PAD,RaceW
```

```
set.seed(1)
```

```
folds=cut(seq(1,nrow(Data2)),breaks=j,labels=F)
```

```
pred=0
```

```
logLike=0
```

```
for (i in 1:j)
```

```
{
```

```
  testIndexes=which(folds==i,arr.ind=T)
```

```
  testData=Data2[testIndexes,]
```

```
  trainData=Data2[-testIndexes,]
```

```
  model =
```

```
  glm(Anastamotic.Leak~Gender+BMI+Age+Race+Tobacco+DM+CAD.PAD+Cancer+Albumin..g.dL.+Operati  
  ve.Length,data=trainData,family="binomial")
```

```
  pred=predict.glm(model,newdata=testData,type="response")
```

```
  logLike=logLike + sum((testData$Anastamotic.Leak*log(pred))+((1-testData$Anastamotic.Leak)*log(1-  
  pred)) )
```

```
}
```

```
print(logLike)
```

```
#drops: Cancer
```

```
set.seed(1)
```

```
folds=cut(seq(1,nrow(Data2)),breaks=j,labels=F)
```

```
pred=0
```

```

logLike=0
for (i in 1:j)
{
  testIndexes=which(folds==i,arr.ind=T)
  testData=Data2[testIndexes,]
  trainData=Data2[-testIndexes,]

  model =
glm(Anastamotic.Leak~Gender+BMI+Age+Race+Tobacco+DM+CAD.PAD+Albumin..g.dL.+Operative.Leng
th,data=trainData,family="binomial")

  pred=predict.glm(model,newdata=testData,type="response")

  logLike=logLike + sum((testData$Anastamotic.Leak*log(pred))+((1-testData$Anastamotic.Leak)*log(1-
pred)) )
}
print(logLike)

```

```

#Potential drops: Cancer,CAD.PAD

set.seed(1)

folds=cut(seq(1,nrow(Data2)),breaks=j,labels=F)

pred=0
logLike=0
for (i in 1:j)
{
  testIndexes=which(folds==i,arr.ind=T)
  testData=Data2[testIndexes,]
  trainData=Data2[-testIndexes,]

  model =
glm(Anastamotic.Leak~Gender+BMI+Age+Race+Tobacco+DM+Albumin..g.dL.+Operative.Length,data=tr
ainData,family="binomial")

  pred=predict.glm(model,newdata=testData,type="response")

  logLike=logLike + sum((testData$Anastamotic.Leak*log(pred))+((1-testData$Anastamotic.Leak)*log(1-
pred)) )
}

```

```

}

print(logLike)

#dropping RaceW does not make sense since its only one of a catagory column

model =
glm(Anastamotic.Leak~Gender+BMI+Age+Race+Tobacco+DM+Albumin..g.dL.+Operative.Length,data=trainData,family="binomial")

modSum=summary(model)

printTable=function()
{
  print("Variable:      Estimate:  STD Error:  Z-Value  Pr(>|z|)   %Change      Lower Bound
Upper Bound")

  print(paste0("Intercept"," ",modSum$coefficients[1,1]," ",modSum$coefficients[1,2],"
",modSum$coefficients[1,3]
      ," ",modSum$coefficients[1,4]," ",round((exp(modSum$coefficients[1,1])-1)*100,8)," ",
      round((exp(modSum$coefficients[1,1]-1.96*modSum$coefficients[1,2])-1)*100,8)," ",
      round(((exp(modSum$coefficients[1,1]+1.96*modSum$coefficients[1,2])-1)*100,8))))
  for(i in 1:8)
  {
    print(paste0(str_pad(attr(model$terms , "term.labels")[i],17,"right"),"
",round(modSum$coefficients[i+1,1],6)," ",round(modSum$coefficients[i+1,2],6)," ",
      round(modSum$coefficients[i+1,3],6)," ",round(modSum$coefficients[i+1,4],6),"
",round((exp(modSum$coefficients[i+1,1])-1)*100,8)," ",
      round((exp(modSum$coefficients[i+1,1]-1.96*modSum$coefficients[i+1,2])-1)*100,8)," ",
      round(((exp(modSum$coefficients[i+1,1]+1.96*modSum$coefficients[i+1,2])-1)*100,8))))
  }
}

```

printTable()