

Problem1

Paul Mendoza

December 9, 2016

```
require(magrittr)
require(dplyr)
require(ggplot2)
require(glmnet)
```

Fit the data in Table 1 to a linear model using:

Table1

```
##      x1    x2    y
## 1  0.99  0.98  6.42
## 2 -0.75 -0.76  0.20
## 3 -0.50 -0.48  0.80
## 4 -1.08 -1.08 -0.57
## 5  0.09  0.09  4.75
## 6 -1.28 -1.27 -1.42
## 7 -0.79 -0.79  1.07
## 8 -1.17 -1.17  0.20
## 9 -0.57 -0.57  1.08
## 10 -1.62 -1.62 -0.15
## 11  0.34  0.35  2.90
## 12  0.51  0.51  3.37
## 13 -0.91 -0.92  0.05
## 14  1.85  1.86  5.50
## 15 -1.12 -1.12  0.17
## 16 -0.70 -0.70  1.72
## 17  1.19  1.18  3.97
## 18  1.24  1.23  6.38
## 19 -0.52 -0.52  3.29
## 20 -1.41 -1.44 -1.49
```

1. Least Squares

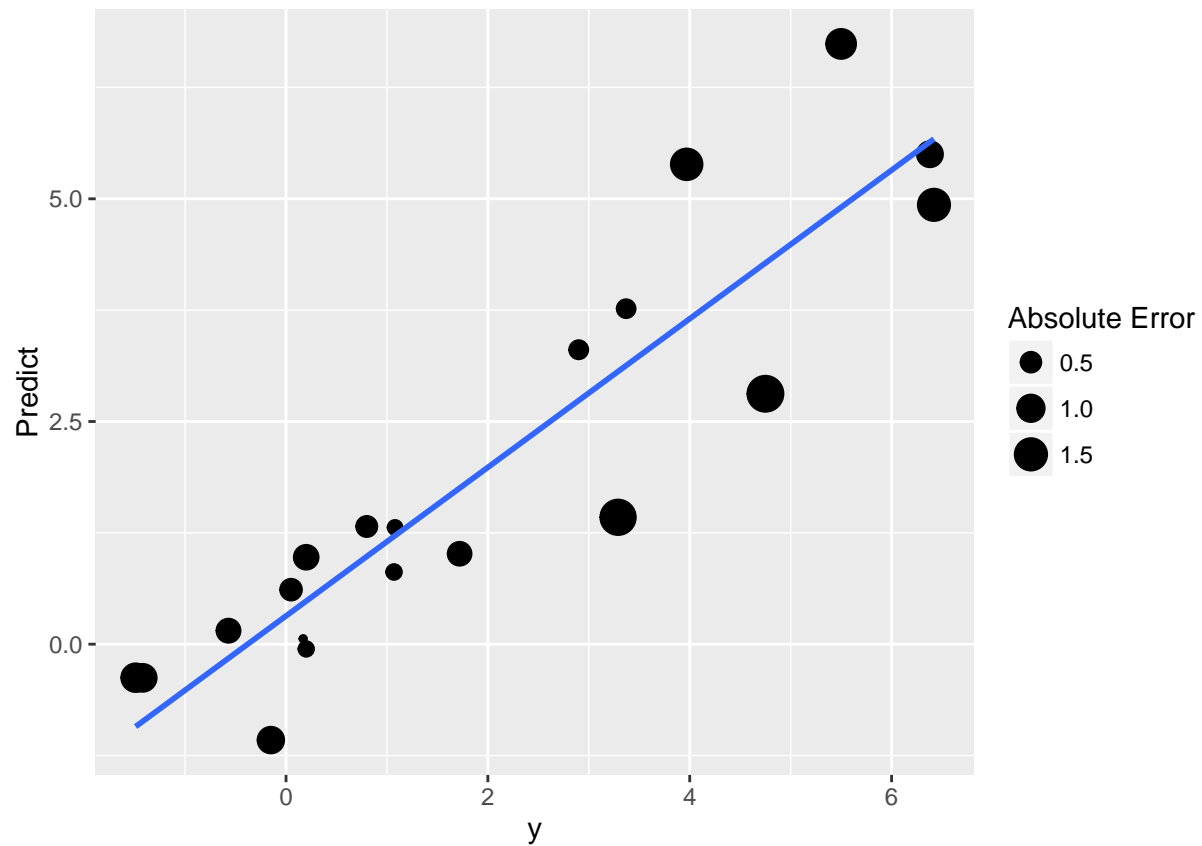
```
LeastFit<-lm(formula = y~x1+x2,data=Table1)
LeastFit
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = Table1)
##
## Coefficients:
## (Intercept)          x1          x2
##      2.607      9.707     -7.433
```

```

plotDF<-Table1
plotDF[, 'Type']<-'Train'
plotDF$Predict<-predict(LeastFit,plotDF[,1:2])
plotDF$Error<-plotDF$y-plotDF$Predict
ggplot(plotDF,aes(x=y,y=Predict,size=abs(Error)))+geom_point()+
  scale_size("Absolute Error")+geom_smooth(method="lm",se=F,size=1)

```



```

sqrt(var(data.frame(plotDF %>% filter(Type=="Train") %>% select(Error))))/20

```

```

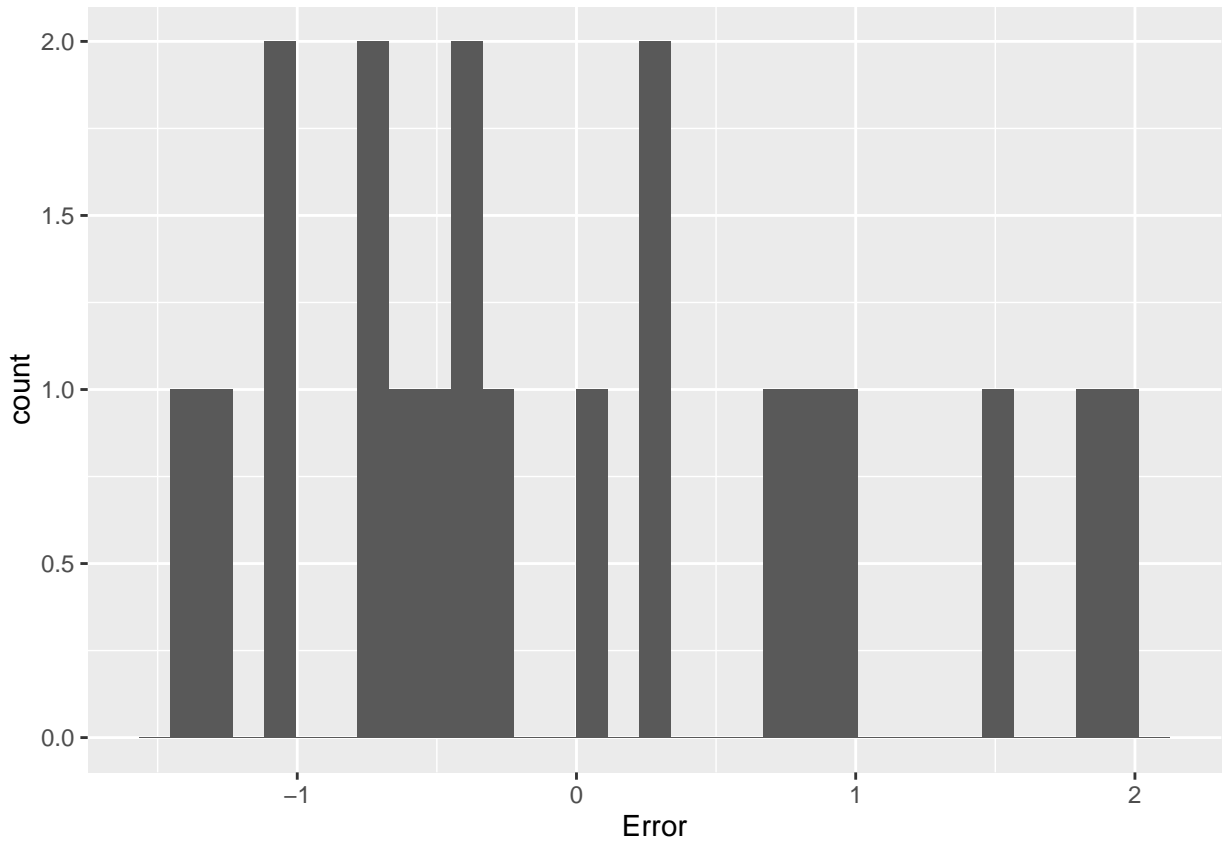
##          Error
## Error 0.05088412

```

```

ggplot(plotDF,aes(x=Error)) + geom_histogram()

```



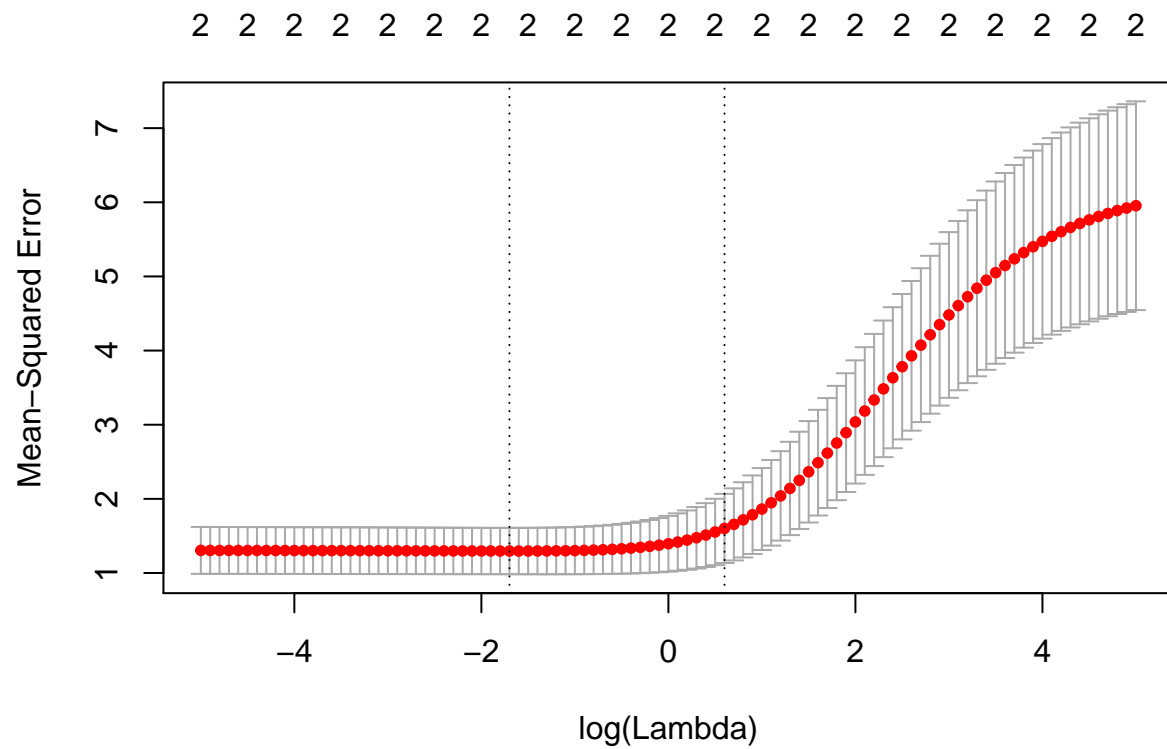
```
sensitivities <- coef(LeastFit)
sensDF <- data.frame(Method = 0, Var = 0, Value=0)
sensDF[1:length(sensitivities), 'Method'] <- "Least-Squares"
sensDF[1:length(sensitivities), 'Var'] <- names(sensitivities)
sensDF[1:length(sensitivities), 'Value'] <- (sensitivities)
rowStart <- length(sensitivities)+1
```

2. Ridge Regression

Ridge regression sets $\alpha=0$, which adds damping to the coefficients

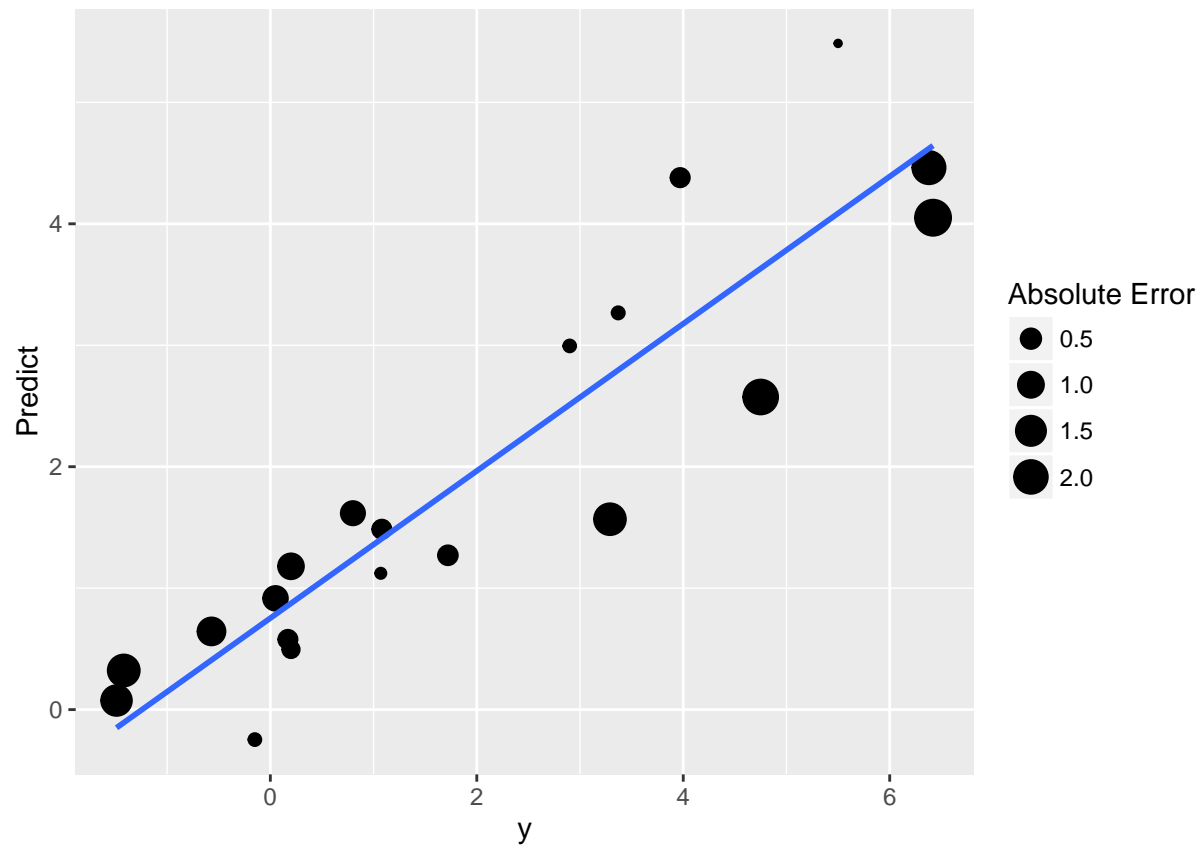
```
crossValid <- cv.glmnet(as.matrix(Table1[,1:2]),
                        as.matrix(Table1$y), alpha = 0,
                        lambda=exp(seq(-5,5,by=0.1)))

plot(crossValid)
```



```
lambda <- crossValid$lambda.min
sensitivities <- coef(crossValid)
plotDF <- Table1
plotDF[, 'Type'] <- 'Train'
plotDF[, "Predict" ]<- data.frame(Predict=predict(crossValid,as.matrix(plotDF[,1:2]),
                                                lambda=lambda))

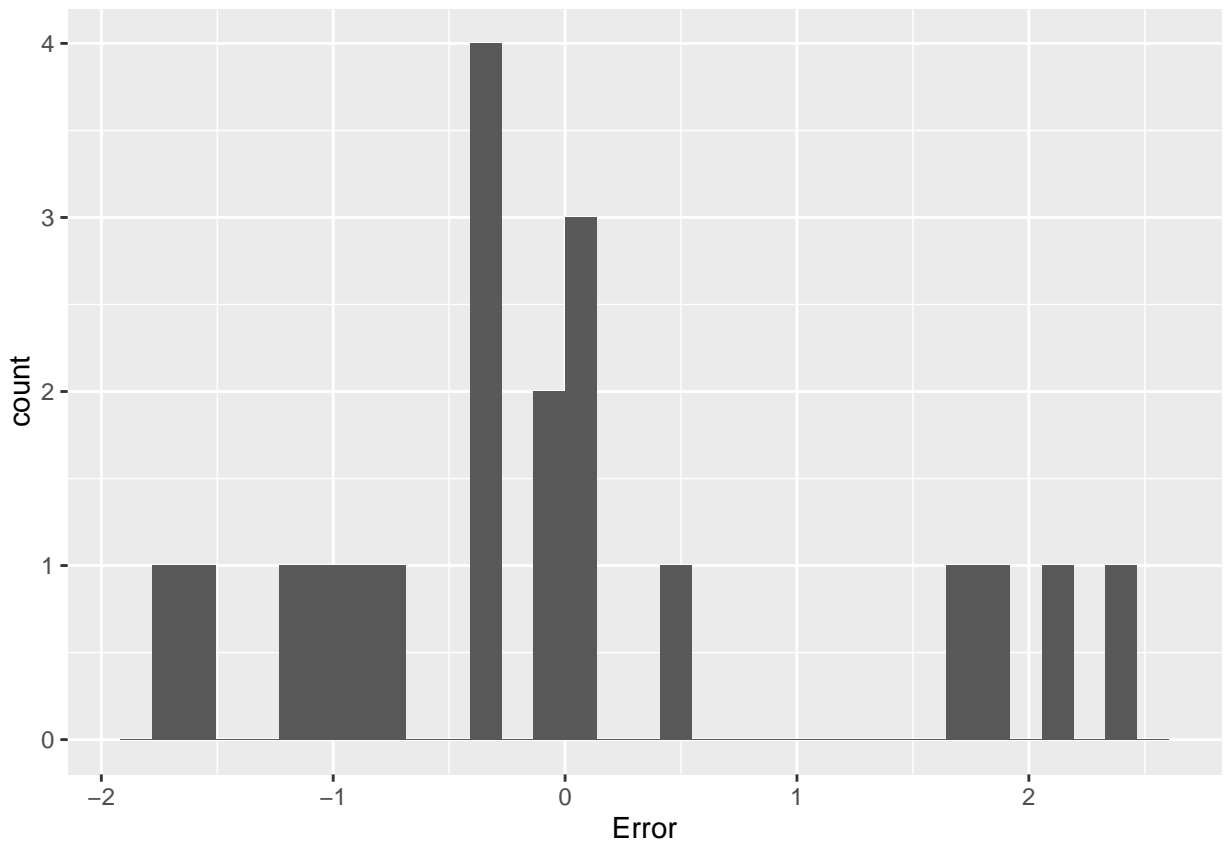
plotDF$Error <- plotDF$y-plotDF$Predict
ggplot(plotDF,aes(x=y,y=Predict,size=abs(Error))) + geom_point() +
scale_size("Absolute Error") + geom_smooth(method="lm",se=F,size=1)
```



```
sqr( var( data.frame( plotDF %>% filter( Type=="Train" ) %>% select( Error ) ) ) ) / 20
```

```
##          Error
## Error 0.05976789
```

```
ggplot( plotDF, aes( x=Error ) ) + geom_histogram()
```

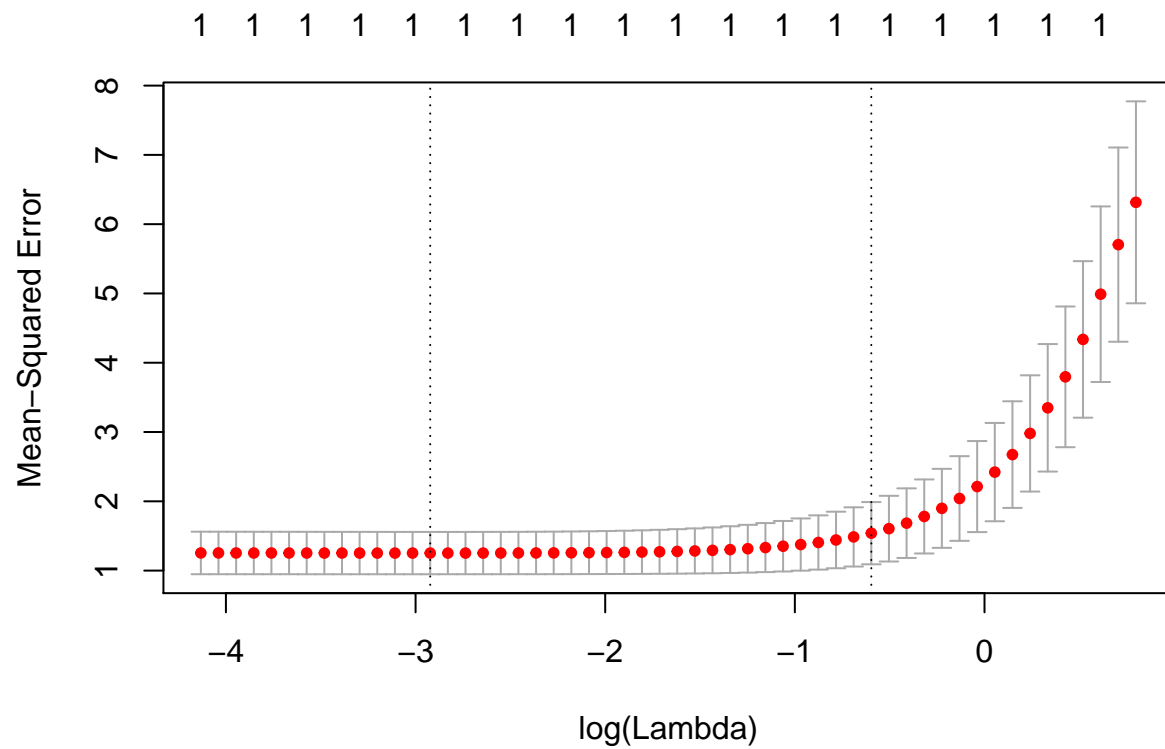


```
sensDF[rowStart:(rowStart + length(sensitivities)-1), 'Method'] <- "Ridge"
sensDF[rowStart:(rowStart+length(sensitivities)-1), 'Var']<-t(t(rownames(sensitivities)))
sensDF[rowStart:(rowStart +length(sensitivities)-1), 'Value']<-as.numeric(sensitivities)
rowStart <- rowStart + length(sensitivities)
```

3. Lasso Regression

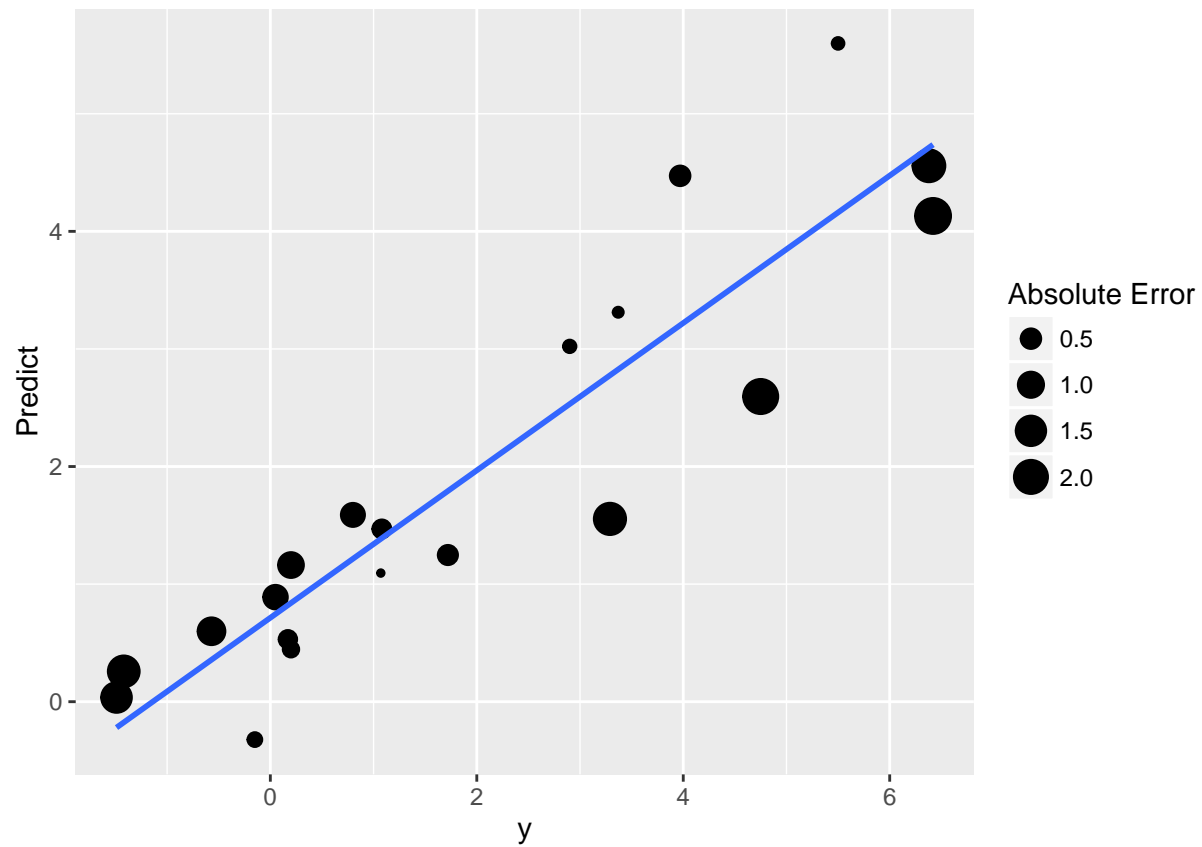
Lasso regression sets alpha=1

```
crossValid <- cv.glmnet(as.matrix(Table1[,1:2]),as.matrix(Table1$y),alpha = 1)
plot(crossValid)
```



```
lambda <- crossValid$lambda.min
sensitivities <- coef(crossValid)
plotDF <- Table1
plotDF[, 'Type'] <- 'Train'
plotDF[, "Predict" ]<- data.frame(Predict=predict(crossValid,as.matrix(Table1[,1:2]),
                                                lambda=lambda))

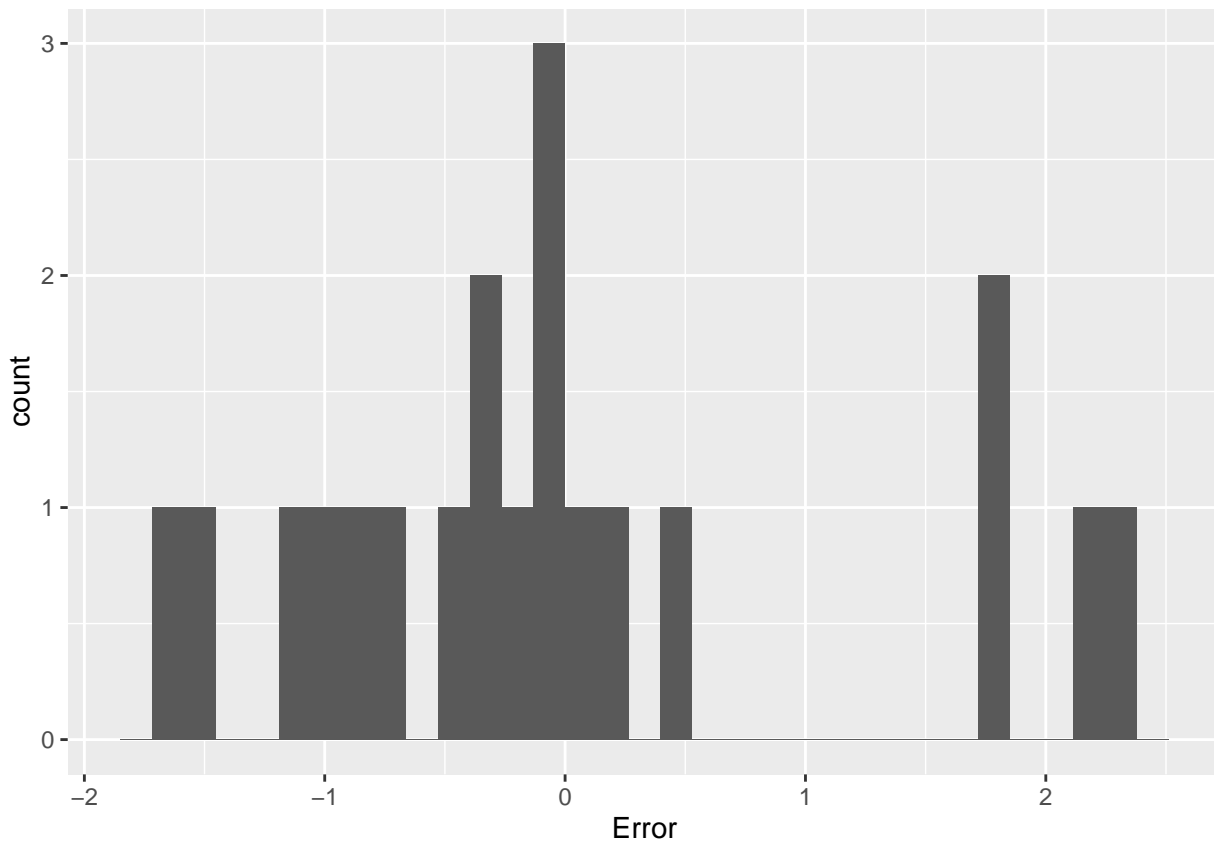
plotDF$Error <- plotDF$y-plotDF$Predict
ggplot(plotDF,aes(x=y,y=Predict,size=abs(Error))) + geom_point() +
scale_size("Absolute Error") + geom_smooth(method="lm",se=F,size=1)
```



```
sqr t(var(data.frame(plotDF %>% filter(Type=="Train") %>% select(Error))))/20
```

```
##          Error
## Error 0.05832321
```

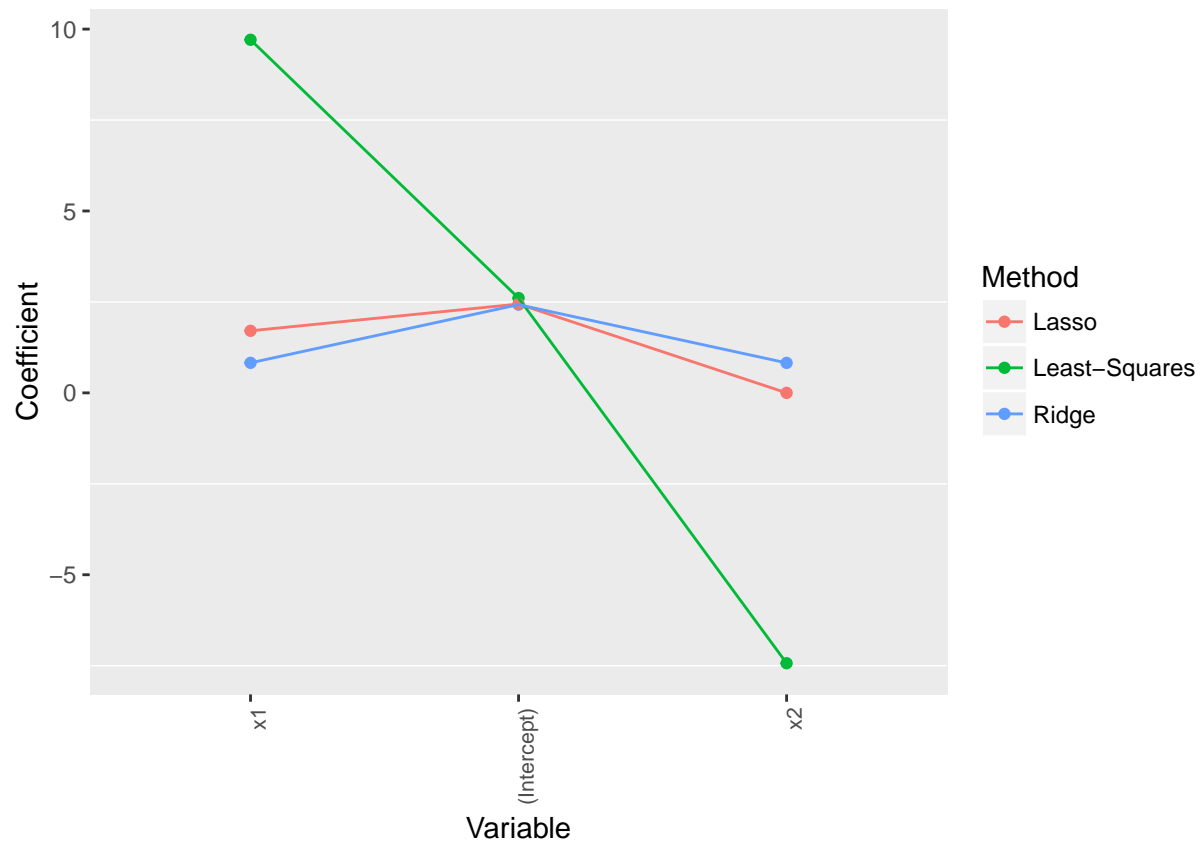
```
ggplot(plotDF,aes(x=Error)) + geom_histogram()
```

```
sensDF[rowStart:(rowStart + length(sensitivities)-1), 'Method'] <- "Lasso"
sensDF[rowStart:(rowStart+length(sensitivities)-1), 'Var']<-t(t(rownames(sensitivities)))
sensDF[rowStart:(rowStart +length(sensitivities)-1), 'Value']<-as.numeric(sensitivities)
rowStart <- rowStart + length(sensitivities)
```

Compare Methods

```
ggplot(sensDF, aes(x=reorder(Var, -Value), y=Value, color=Method, group=Method)) +
  geom_point() + geom_line() +
  theme(panel.grid.major = element_blank(),
        axis.text.x = element_text(angle = 90, hjust = 1, size=8))+
  scale_x_discrete("Variable") + scale_y_continuous("Coefficient")
```



The Lasso and Ridge are both more bounded in their coefficients.