Modeling

Michael Bostwick

2/7/2018

Modeling

Variable pre-selection

Check for highly correlated variables. Variables that have perfect or very high correlation will have one of the pair removed (marked as a “C” in the Variables spreadsheet).

```
##                                       Var1        Var2     value
## 16319                                  t10          t8 0.9066941
## 25385 produccion_en_libras_producto_vendido        ad11 0.9381219
## 31087                                a24_f       a16_b 0.9130902
## 44556                               tp52_g      tp48_g 0.9778341
## 44820                               to57_e      to53_e 0.9143210
## 48441                            percperm2    percperm 0.9274772
## 49477                         percpasture2 percpasture 0.9737279
## 49996                           percother2   percother 0.9465343
```

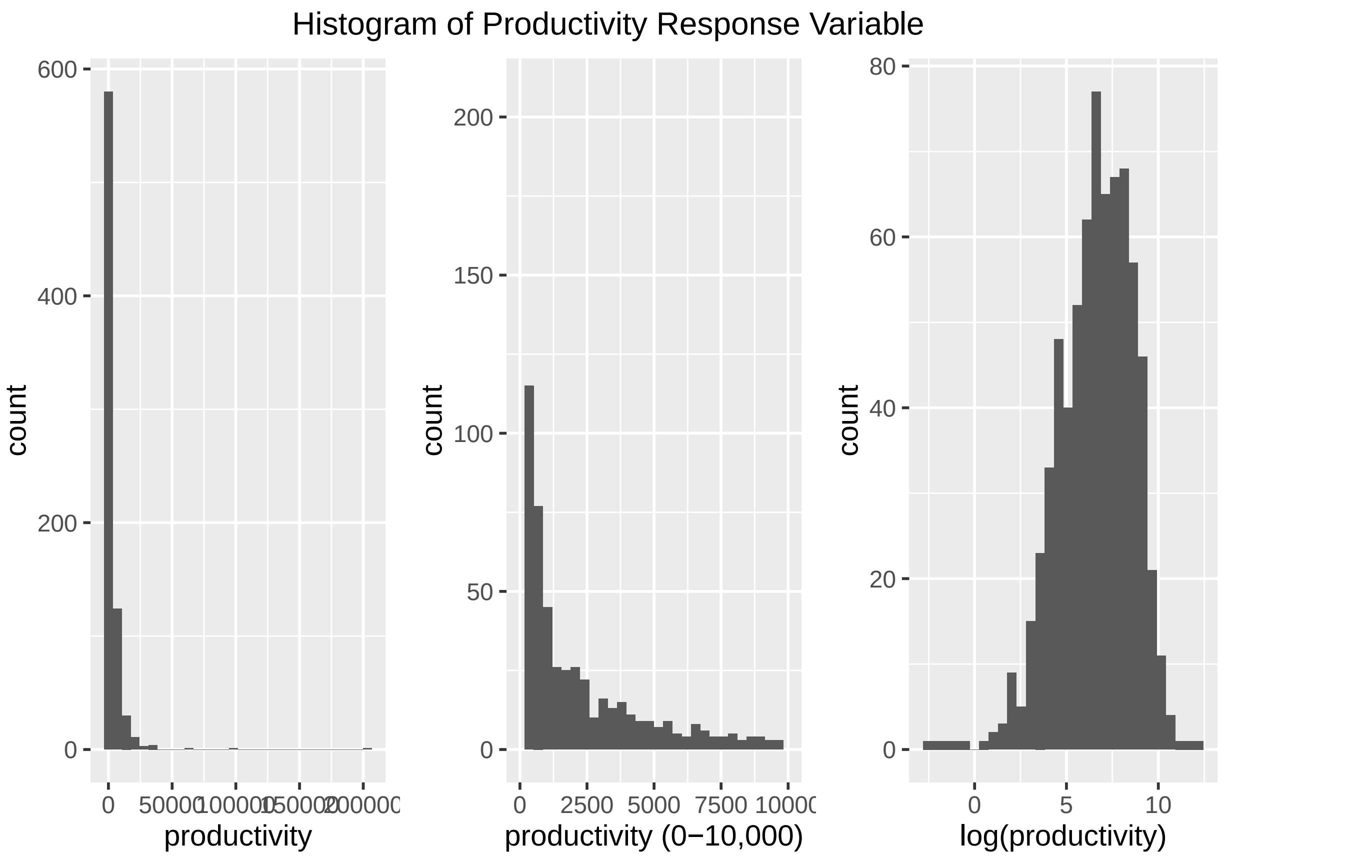
UPA Production

```
zero_prod <- reduced_data[reduced_data$productivity ==0,]
summary(zero_prod$netincome)
```

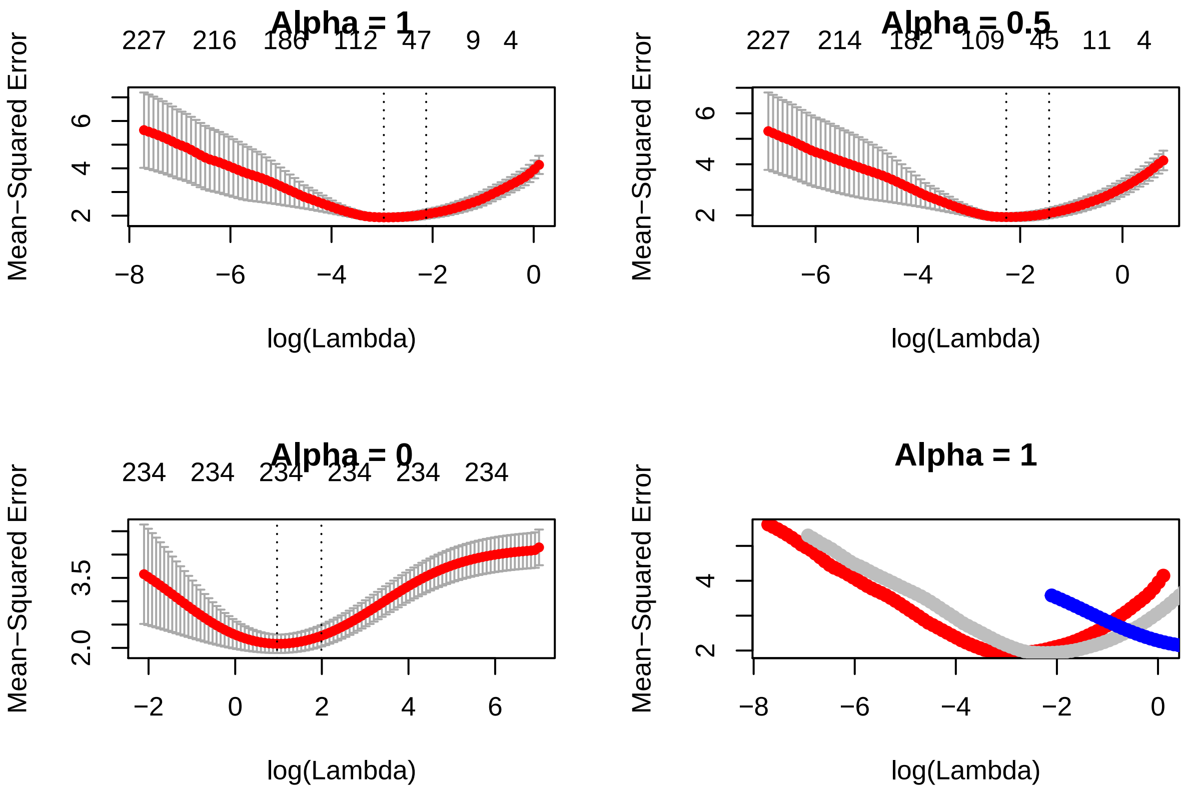
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -84380   -7020   -1490   -8633    -283    9399
```

```
par(mfrow = c(2,1))
raw_prod_hist <- ggplot(data = reduced_data) + geom_histogram(mapping = aes(x = productivity))
log_prod_hist <- ggplot(data = reduced_data) + geom_histogram(mapping = aes(x = log(productivity)))
lower_prod_hist <- ggplot(data = reduced_data) + geom_histogram(mapping = aes(x = productivity)) +
  xlim(0, 10000) + xlab("productivity (0-10,000)")
grid.arrange(raw_prod_hist, lower_prod_hist, log_prod_hist, nrow = 1, top = "Histogram of Productivity I
```

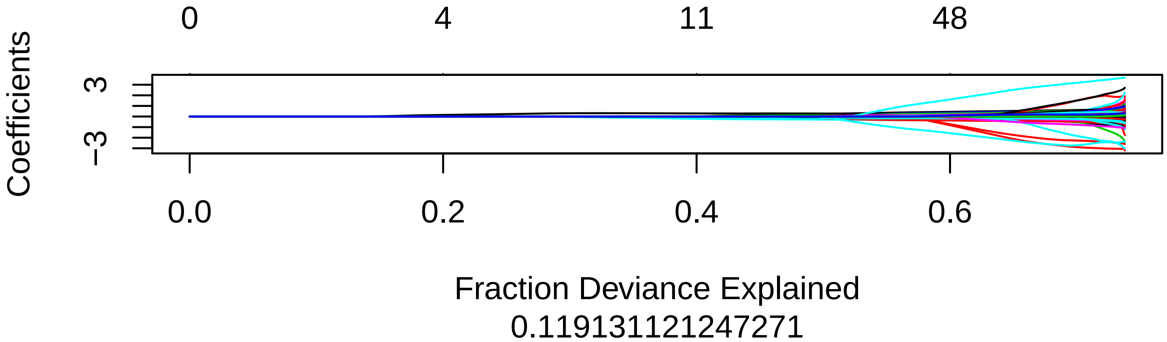
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 56 rows containing non-finite values (stat_bin).

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 38 rows containing non-finite values (stat_bin).
```

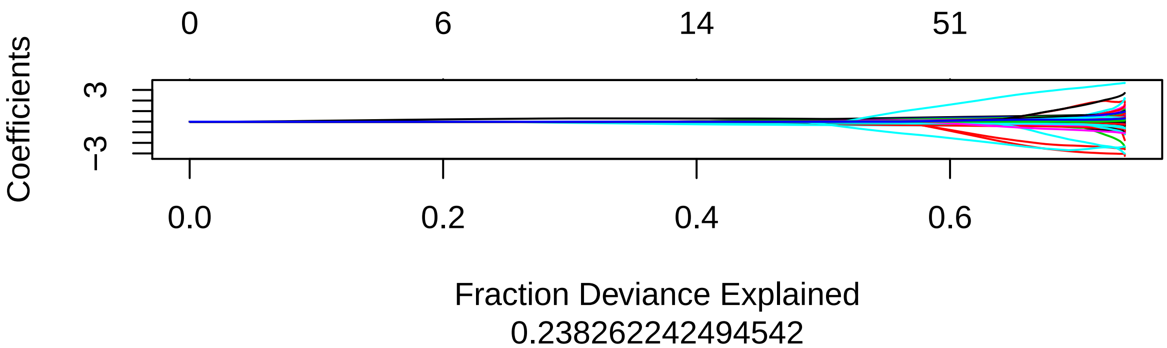



```r
production_include <- vars[which(is.na(vars$`UPA Production`)),1]$`Variable Name`
production_x <- subset(reduced_data[reduced_data$productivity > 0,], select = production_include)
production_x <- model.matrix(~., production_x)[,-1]

log_productivity <- log(reduced_data[reduced_data$productivity > 0, 'productivity'])

foldid=sample(1:10,size=length(log_productivity),replace=TRUE)
cv1=cv.glmnet(production_x,log_productivity,foldid=foldid,alpha=1)
cv.5=cv.glmnet(production_x,log_productivity,foldid=foldid,alpha=.5)
cv0=cv.glmnet(production_x,log_productivity,foldid=foldid,alpha=0)

par(mfrow=c(2,2))
plot(cv1,main="Alpha = 1");plot(cv.5,main="Alpha = 0.5");plot(cv0,main="Alpha = 0");
plot(log(cv1$lambda),cv1$cvm,pch=19,col="red",xlab="log(Lambda)",ylab=cv1$name,  main="Alpha = 1")
points(log(cv.5$lambda),cv.5$cvm,pch=19,col="grey", main="Alpha = 0.5")
points(log(cv0$lambda),cv0$cvm,pch=19,col="blue", main="Alpha = 0")
```




```r
par(mfrow=c(1,1))

par(mfrow=c(2,1))
plot(cv1$glmnet.fit, s = cv1$lambda.1se, xvar = "dev", label = FALSE)
plot(cv.5$glmnet.fit, s = cv.5$lambda.1se, xvar = "dev", label = FALSE)
```






```
lasso.coef <- predict(cv1,type="coefficients",s=cv1$lambda.1se)[1:227,]
lasso.coef[lasso.coef!=0]
```

```
##                       (Intercept)
##                      6.209901e+00
##         `CPermanentes_OTROS BANANOS`
##                      3.878382e-01
##                 CPermanentes_PLATANO
##                      2.893157e-01
##         `CTransitorios_TOMATE RINON`
##                      2.507306e-01
##                   CTransitorios_YUCA
##                      1.857631e-02
##                          num_cultivo
##                      1.529339e-05
##                    Pastos_BRACHIARIA
##                     -3.499138e-01
##                      Pastos_ELEFANTE
##                     -1.473814e-01
##                  `Pastos_KING GRASS`
##                     -4.483688e-02
##                             pc4None
##                      1.223847e-01
##                                 pc6
##                     -3.876450e-03
##                       Arboles_GUABA
##                      1.515741e-01
##                      Arboles_GUINEO
##                      1.136355e-01
##           `Arboles_LIMON MANDARINA`
```

```
##                                     9.626446e-02
##                                  Arboles_NARANJA
##                                     5.683590e-03
##                                             ad11
##                                     3.213144e-05
## produccion_en_libras_producto_vendido
##                                     1.798939e-05
##                                            v30_a
##                                    -2.997173e-03
##                                              c12
##                                     2.018429e-06
##                                            ga9Si
##                                     5.476493e-02
##                        ga15_cualADECUACION UPA
##                                    -1.711513e-02
##                               ga15_cualPACHETE
##                                     1.028400e+00
##                                         percperm
##                                     1.439874e-02
##                                         perctemp
##                                     1.857261e-02
##                                       percfallow
##                                     9.513629e-05
##                                      percpasture
##                                    -6.248586e-03
##                                        percbrush
##                                    -1.020253e-03
##                                        percother
##                                     1.646750e-02
##                                        perctemp2
##                                     1.147495e-03
##                                         percinv2
##                                    -1.209785e-03
##                                             d3Si
##                                    -4.102828e-02
##                            ReclassCONSERVACION
##                                    -3.415214e-01
##                                 ReclassPECUARIO
##                                    -1.772193e-01
##                                   ABANDONEDTRUE
##                                    -1.378821e-01
##                                CONSERVATIONTRUE
##                                    -5.645474e-02
##                                    FORESTRYTRUE
##                                    -1.399897e-01
##                                     LODGINGTRUE
##                                    -2.307416e-02
##      `ENERGIA_ELENERGIA SOLAR PRIVADA`
##                                    -1.152407e+00
```