Homework 9

Megan Marziali

3/17/2021

# Question 1

## Loading and preparing data

Loading data into single data frame.

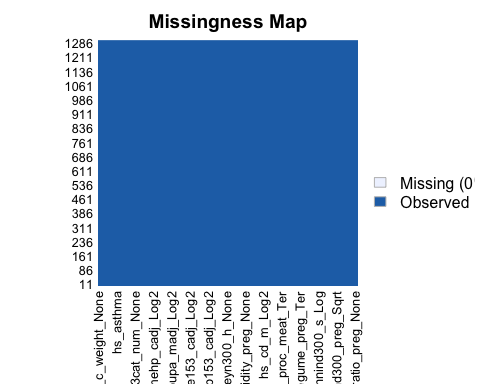
library(tidyverse)  
library(caret)  
library(rpart.plot)  
library(Amelia)  
library(arsenal)  
  
#Load data using path of where file is stored  
load("./data/exposome.RData")  
  
#Merge all data frames into a single data frame. FYI, this is just a shortcut by combining baseR with piping from tidyverse. There are other ways of merging across three data frames that are likely more elegant.  
  
studydata =   
 merge(exposome,phenotype,by = "ID") %>%   
 merge(covariates, by = "ID")  
  
#Strip off ID Variable  
studydata$ID = NULL

Data cleaning, and removing non-modifiable variables from the possible features.

data.rec = studydata %>%   
 janitor::clean\_names() %>%   
 mutate(  
 hs\_asthma = as.factor(hs\_asthma)  
 ) %>%   
 select(-h\_accesslines300\_preg\_dic0,  
 -h\_accesspoints300\_preg\_log,  
 -h\_builtdens300\_preg\_sqrt,  
 -h\_connind300\_preg\_sqrt,  
 -h\_fdensity300\_preg\_log,  
 -h\_frichness300\_preg\_none,  
 -h\_landuseshan300\_preg\_none,  
 -h\_popdens\_preg\_sqrt,  
 -hs\_accesslines300\_h\_dic0,  
 -h\_walkability\_mean\_preg\_none,  
 -hs\_accesspoints300\_h\_log,  
 -hs\_builtdens300\_h\_sqrt,  
 -hs\_connind300\_h\_log,  
 -hs\_fdensity300\_h\_log,  
 -hs\_landuseshan300\_h\_none,  
 -hs\_popdens\_h\_sqrt,  
 -hs\_walkability\_mean\_h\_none,  
 -hs\_accesslines300\_s\_dic0,  
 -hs\_accesspoints300\_s\_log,  
 -hs\_builtdens300\_s\_sqrt,  
 -hs\_connind300\_s\_log,  
 -hs\_fdensity300\_s\_log,  
 -hs\_landuseshan300\_s\_none,  
 -hs\_popdens\_s\_sqrt,  
 -h\_humidity\_preg\_none,  
 -h\_pressure\_preg\_none,  
 -h\_temperature\_preg\_none,  
 -hs\_hum\_mt\_hs\_h\_none,  
 -hs\_tm\_mt\_hs\_h\_none,  
 -hs\_uvdvf\_mt\_hs\_h\_none,  
 -hs\_hum\_dy\_hs\_h\_none,  
 -hs\_hum\_wk\_hs\_h\_none,  
 -hs\_tm\_dy\_hs\_h\_none,  
 -hs\_tm\_wk\_hs\_h\_none,  
 -hs\_uvdvf\_dy\_hs\_h\_none,  
 -hs\_uvdvf\_wk\_hs\_h\_none,  
 -hs\_blueyn300\_s\_none,  
 -h\_blueyn300\_preg\_none,  
 -h\_greenyn300\_preg\_none,  
 -h\_ndvi100\_preg\_none,  
 -hs\_greenyn300\_s\_none,  
 -hs\_blueyn300\_h\_none,  
 -hs\_greenyn300\_h\_none,  
 -hs\_ndvi100\_h\_none,  
 -hs\_ndvi100\_s\_none,  
 -h\_lden\_cat\_preg\_none,  
 -hs\_ln\_cat\_h\_none,  
 -hs\_lden\_cat\_s\_none,  
 -h\_distinvnear1\_preg\_log,  
 -h\_trafload\_preg\_pow1over3,  
 -h\_trafnear\_preg\_pow1over3,  
 -hs\_trafload\_h\_pow1over3,  
 -hs\_trafnear\_h\_pow1over3,  
 -hs\_gen\_tot,  
 -hs\_c\_height\_none,  
 -hs\_c\_weight\_none,  
 -e3\_gac\_none,  
 -e3\_sex\_none,  
 -e3\_yearbir\_none,  
 -h\_age\_none,  
 -h\_cohort,  
 -h\_parity\_none,  
 -hs\_child\_age\_none,  
 -hs\_zbmi\_who,  
 -hs\_correct\_raven,  
 -hs\_bmi\_c\_cat)

Data exploration.

# Investigating missing data  
missmap(studydata)



# No missingness observed.  
  
# Exploring continuous/categorical variables of interest  
table.1 = tableby(~ hs\_asthma + hs\_no2\_wk\_hs\_h\_log + h\_pm\_log +  
 h\_folic\_t1\_none + fas\_cat\_none + hs\_hm\_pers\_none + hs\_participation\_3cat\_none,  
 data = data.rec,  
 numeric.stats = c("mean","median", "range"))  
summary(table.1, text = TRUE)

|  |  |
| --- | --- |
|  | Overall (N=1301) |
| hs\_asthma |  |
| - 0 | 1159 (89.1%) |
| - 1 | 142 (10.9%) |
| hs\_no2\_wk\_hs\_h\_log |  |
| - Mean | 2.864 |
| - Median | 2.981 |
| - Range | 0.952 - 4.805 |
| h\_pm\_log |  |
| - Mean | 2.443 |
| - Median | 2.304 |
| - Range | 1.549 - 5.236 |
| h\_folic\_t1\_none |  |
| - 0 | 606 (46.6%) |
| - 1 | 695 (53.4%) |
| fas\_cat\_none |  |
| - Low | 146 (11.2%) |
| - Middle | 486 (37.4%) |
| - High | 669 (51.4%) |
| hs\_hm\_pers\_none |  |
| - Mean | 4.248 |
| - Median | 4.000 |
| - Range | 1.000 - 10.000 |
| hs\_participation\_3cat\_none |  |
| - None | 748 (57.5%) |
| - 1 organisation | 355 (27.3%) |
| - 2 or more organisations | 198 (15.2%) |

I chose to explore a few variables related to asthma and air pollution, including NO2 concentration (mean: 2.86, median: 2.98; range: 0.95 - 4.81). I also opted to explore concentration of particulate matter (mean: 2.44; median: 2.30; range: 1.55 - 5.24). Within this sample, 53.4% (N=695) of mothers consumed folic acid during pregnancy. 51.4% (N=669) of mothers scored high on the family affluence score. The median number of people leaving in the home was 4 (range: 1-10). The majority (57.5%, N=748) of mothers did not participate in any organizations.

Partitioning data.

set.seed(100)  
  
#Partition data for use in demonstration  
train.indices = createDataPartition(y = data.rec$hs\_asthma, p = 0.7,list = FALSE)  
training = data.rec[train.indices, ]  
testing = data.rec[-train.indices, ]

# Question 2

## Developing research question

Given available data, and the number of features, my research question of interest will be hypothesis generating. This analysis will aim to determine potential risk factors of asthma at 6-11 years old.

# Question 3

## Using LASSO for feature selection

set.seed(100)  
  
#Create grid to search lambda  
lambda = 10^seq(-3,3, length = 100)  
  
lasso.m = train(  
 hs\_asthma ~.,   
 data = training,   
 method = "glmnet",   
 trControl = trainControl("cv", number = 10, sampling = "down"),   
 tuneGrid = expand.grid(alpha = 0, lambda = lambda)  
)  
  
#Print the values of alpha and lambda that gave best prediction  
lasso.m$bestTune

## alpha lambda  
## 56 0 2.154435

#Print all of the options examined  
lasso.m$results

## alpha lambda Accuracy Kappa AccuracySD KappaSD  
## 1 0 1.000000e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 2 0 1.149757e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 3 0 1.321941e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 4 0 1.519911e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 5 0 1.747528e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 6 0 2.009233e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 7 0 2.310130e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 8 0 2.656088e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 9 0 3.053856e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 10 0 3.511192e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 11 0 4.037017e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 12 0 4.641589e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 13 0 5.336699e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 14 0 6.135907e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 15 0 7.054802e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 16 0 8.111308e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 17 0 9.326033e-03 0.5494505 0.038620201 0.08102834 0.09286901  
## 18 0 1.072267e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 19 0 1.232847e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 20 0 1.417474e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 21 0 1.629751e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 22 0 1.873817e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 23 0 2.154435e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 24 0 2.477076e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 25 0 2.848036e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 26 0 3.274549e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 27 0 3.764936e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 28 0 4.328761e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 29 0 4.977024e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 30 0 5.722368e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 31 0 6.579332e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 32 0 7.564633e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 33 0 8.697490e-02 0.5494505 0.038620201 0.08102834 0.09286901  
## 34 0 1.000000e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 35 0 1.149757e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 36 0 1.321941e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 37 0 1.519911e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 38 0 1.747528e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 39 0 2.009233e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 40 0 2.310130e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 41 0 2.656088e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 42 0 3.053856e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 43 0 3.511192e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 44 0 4.037017e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 45 0 4.641589e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 46 0 5.336699e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 47 0 6.135907e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 48 0 7.054802e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 49 0 8.111308e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 50 0 9.326033e-01 0.5494505 0.038620201 0.08102834 0.09286901  
## 51 0 1.072267e+00 0.5505375 0.039238874 0.08068749 0.09337532  
## 52 0 1.232847e+00 0.5494505 0.038834891 0.08299165 0.09312536  
## 53 0 1.417474e+00 0.5516484 0.045959232 0.07965892 0.09200609  
## 54 0 1.629751e+00 0.5527353 0.046598958 0.08028776 0.09168500  
## 55 0 1.873817e+00 0.5549331 0.045100504 0.08131766 0.09102755  
## 56 0 2.154435e+00 0.5560201 0.045508635 0.08140654 0.09106500  
## 57 0 2.477076e+00 0.5549331 0.045464291 0.08391618 0.09199355  
## 58 0 2.848036e+00 0.5527234 0.040627637 0.08396003 0.09155058  
## 59 0 3.274549e+00 0.5527353 0.041036402 0.08544805 0.09206372  
## 60 0 3.764936e+00 0.5527353 0.039950533 0.08305927 0.09106914  
## 61 0 4.328761e+00 0.5494505 0.038047233 0.08015269 0.09032305  
## 62 0 4.977024e+00 0.5461300 0.035325006 0.07764791 0.08812400  
## 63 0 5.722368e+00 0.5494147 0.037427834 0.07925697 0.08987445  
## 64 0 6.579332e+00 0.5494147 0.037868814 0.08289778 0.09092903  
## 65 0 7.564633e+00 0.5493908 0.037689543 0.08156605 0.09114023  
## 66 0 8.697490e+00 0.5493908 0.036970834 0.08222141 0.09272925  
## 67 0 1.000000e+01 0.5482919 0.032676651 0.08114459 0.08818423  
## 68 0 1.149757e+01 0.5504778 0.042007046 0.07786951 0.07426965  
## 69 0 1.321941e+01 0.5504778 0.045753582 0.07735085 0.07198450  
## 70 0 1.519911e+01 0.5504778 0.045879885 0.07786951 0.07263089  
## 71 0 1.747528e+01 0.5504778 0.045879885 0.07786951 0.07263089  
## 72 0 2.009233e+01 0.5515886 0.043364154 0.07905731 0.06750314  
## 73 0 2.310130e+01 0.5515767 0.043399590 0.07963403 0.06744855  
## 74 0 2.656088e+01 0.5504778 0.042923921 0.08124262 0.06748494  
## 75 0 3.053856e+01 0.5526756 0.044338665 0.08297691 0.06786715  
## 76 0 3.511192e+01 0.5515647 0.043348558 0.08073207 0.06733710  
## 77 0 4.037017e+01 0.5504658 0.042582719 0.07949984 0.06695619  
## 78 0 4.641589e+01 0.5493669 0.041961639 0.07992778 0.06630291  
## 79 0 5.336699e+01 0.5515528 0.042969524 0.08065561 0.06503828  
## 80 0 6.135907e+01 0.5515528 0.042969524 0.08065561 0.06503828  
## 81 0 7.054802e+01 0.5515528 0.042969524 0.08065561 0.06503828  
## 82 0 8.111308e+01 0.5537506 0.044365910 0.08237039 0.06536512  
## 83 0 9.326033e+01 0.5526517 0.043742051 0.08201800 0.06542447  
## 84 0 1.072267e+02 0.4531653 0.025399467 0.19634912 0.06141062  
## 85 0 1.232847e+02 0.3180005 0.008374487 0.28795900 0.05815016  
## 86 0 1.417474e+02 0.3063545 0.010224215 0.33280362 0.03233181  
## 87 0 1.629751e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 88 0 1.873817e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 89 0 2.154435e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 90 0 2.477076e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 91 0 2.848036e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 92 0 3.274549e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 93 0 3.764936e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 94 0 4.328761e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 95 0 4.977024e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 96 0 5.722368e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 97 0 6.579332e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 98 0 7.564633e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 99 0 8.697490e+02 0.2656952 0.000000000 0.32909581 0.00000000  
## 100 0 1.000000e+03 0.2656952 0.000000000 0.32909581 0.00000000

# Model coefficients  
coef(lasso.m$finalModel, lasso.m$bestTune$lambda)

## 212 x 1 sparse Matrix of class "dgCMatrix"  
## 1  
## (Intercept) 1.264104e+00  
## h\_abs\_ratio\_preg\_log -7.596667e-02  
## h\_no2\_ratio\_preg\_log -4.294128e-02  
## h\_pm10\_ratio\_preg\_none -3.031655e-03  
## h\_pm25\_ratio\_preg\_none -3.699791e-03  
## hs\_no2\_dy\_hs\_h\_log -2.401997e-02  
## hs\_no2\_wk\_hs\_h\_log -5.508167e-03  
## hs\_no2\_yr\_hs\_h\_log -1.689974e-02  
## hs\_pm10\_dy\_hs\_h\_none -2.699938e-04  
## hs\_pm10\_wk\_hs\_h\_none -1.079465e-03  
## hs\_pm10\_yr\_hs\_h\_none -2.414826e-03  
## hs\_pm25\_dy\_hs\_h\_none -1.550412e-03  
## hs\_pm25\_wk\_hs\_h\_none -2.659160e-03  
## hs\_pm25\_yr\_hs\_h\_none -1.919201e-03  
## hs\_pm25abs\_dy\_hs\_h\_log -2.510077e-02  
## hs\_pm25abs\_wk\_hs\_h\_log -4.690602e-02  
## hs\_pm25abs\_yr\_hs\_h\_log -1.332630e-01  
## h\_absorbance\_log 1.021260e-02  
## h\_benzene\_log 8.167058e-02  
## h\_no2\_log 6.969341e-03  
## h\_pm\_log 3.459284e-02  
## h\_tex\_log -3.424648e-02  
## e3\_alcpreg\_yn\_none1 1.027634e-02  
## h\_bfdur\_ter(10.8,34.9] 3.530504e-03  
## h\_bfdur\_ter(34.9,Inf] -1.982536e-02  
## h\_cereal\_preg\_ter(9,27.3] -4.691329e-03  
## h\_cereal\_preg\_ter(27.3,Inf] 1.674101e-02  
## h\_dairy\_preg\_ter(17.1,27.1] 1.115995e-02  
## h\_dairy\_preg\_ter(27.1,Inf] 1.248846e-02  
## h\_fastfood\_preg\_ter(0.25,0.83] -5.957803e-02  
## h\_fastfood\_preg\_ter(0.83,Inf] 4.496450e-02  
## h\_fish\_preg\_ter(1.9,4.1] 8.943739e-02  
## h\_fish\_preg\_ter(4.1,Inf] -5.675755e-02  
## h\_folic\_t1\_none1 -9.837973e-03  
## h\_fruit\_preg\_ter(0.6,18.2] 1.524462e-02  
## h\_fruit\_preg\_ter(18.2,Inf] -2.124369e-02  
## h\_legume\_preg\_ter(0.5,2] -3.316973e-02  
## h\_legume\_preg\_ter(2,Inf] 1.420568e-02  
## h\_meat\_preg\_ter(6.5,10] -2.839905e-02  
## h\_meat\_preg\_ter(10,Inf] 1.621503e-02  
## h\_pamod\_t3\_noneOften -1.895197e-02  
## h\_pamod\_t3\_noneSometimes 5.114147e-02  
## h\_pamod\_t3\_noneVery Often 3.547797e-03  
## h\_pavig\_t3\_noneLow -3.672866e-03  
## h\_pavig\_t3\_noneMedium -7.395044e-03  
## h\_veg\_preg\_ter(8.8,16.5] -1.987819e-02  
## h\_veg\_preg\_ter(16.5,Inf] 2.657535e-02  
## hs\_bakery\_prod\_ter(2,6] 2.894343e-02  
## hs\_bakery\_prod\_ter(6,Inf] -6.393109e-03  
## hs\_beverages\_ter(0.132,1] 3.410109e-04  
## hs\_beverages\_ter(1,Inf] 2.639916e-02  
## hs\_break\_cer\_ter(1.1,5.5] 3.241453e-02  
## hs\_break\_cer\_ter(5.5,Inf] -3.999152e-03  
## hs\_caff\_drink\_ter(0.132,Inf] -9.921466e-03  
## hs\_dairy\_ter(14.6,25.6] 9.606679e-03  
## hs\_dairy\_ter(25.6,Inf] 2.659620e-02  
## hs\_fastfood\_ter(0.132,0.5] -5.281661e-02  
## hs\_fastfood\_ter(0.5,Inf] 1.558335e-02  
## hs\_kidmed\_none -1.065638e-02  
## hs\_mvpa\_prd\_alt\_none 3.908201e-05  
## hs\_org\_food\_ter(0.132,1] 1.858861e-02  
## hs\_org\_food\_ter(1,Inf] -2.357578e-02  
## hs\_pet\_cat\_r2\_none1 5.992084e-02  
## hs\_pet\_dog\_r2\_none1 9.058621e-03  
## hs\_pet\_noneYes 2.932485e-02  
## hs\_proc\_meat\_ter(1.5,4] 1.309236e-02  
## hs\_proc\_meat\_ter(4,Inf] 3.659459e-03  
## hs\_readymade\_ter(0.132,0.5] 1.785411e-02  
## hs\_readymade\_ter(0.5,Inf] -3.914497e-02  
## hs\_sd\_wk\_none -6.685896e-05  
## hs\_total\_bread\_ter(7,17.5] -4.466059e-02  
## hs\_total\_bread\_ter(17.5,Inf] 4.164095e-02  
## hs\_total\_cereal\_ter(14.1,23.6] -3.432272e-02  
## hs\_total\_cereal\_ter(23.6,Inf] 2.101749e-02  
## hs\_total\_fish\_ter(1.5,3] 6.265517e-03  
## hs\_total\_fish\_ter(3,Inf] -4.031013e-02  
## hs\_total\_fruits\_ter(7,14.1] -3.645041e-02  
## hs\_total\_fruits\_ter(14.1,Inf] 9.731345e-03  
## hs\_total\_lipids\_ter(3,7] -3.920820e-03  
## hs\_total\_lipids\_ter(7,Inf] 2.284498e-02  
## hs\_total\_meat\_ter(6,9] -3.855042e-02  
## hs\_total\_meat\_ter(9,Inf] 3.588180e-02  
## hs\_total\_potatoes\_ter(3,4] 3.715955e-03  
## hs\_total\_potatoes\_ter(4,Inf] 1.714274e-03  
## hs\_total\_sweets\_ter(4.1,8.5] 2.391390e-02  
## hs\_total\_sweets\_ter(8.5,Inf] -7.306111e-04  
## hs\_total\_veg\_ter(6,8.5] 1.826797e-02  
## hs\_total\_veg\_ter(8.5,Inf] 2.086307e-02  
## hs\_total\_yog\_ter(6,8.5] -2.259280e-02  
## hs\_total\_yog\_ter(8.5,Inf] 2.134372e-02  
## hs\_dif\_hours\_total\_none 1.619989e-02  
## hs\_as\_c\_log2 -1.619780e-03  
## hs\_as\_m\_log2 9.309926e-04  
## hs\_cd\_c\_log2 6.090766e-03  
## hs\_cd\_m\_log2 -2.530843e-02  
## hs\_co\_c\_log2 -2.537839e-02  
## hs\_co\_m\_log2 1.751035e-02  
## hs\_cs\_c\_log2 1.485685e-03  
## hs\_cs\_m\_log2 -1.060699e-02  
## hs\_cu\_c\_log2 1.642433e-03  
## hs\_cu\_m\_log2 -4.548935e-02  
## hs\_hg\_c\_log2 1.028503e-03  
## hs\_hg\_m\_log2 -1.492716e-02  
## hs\_mn\_c\_log2 4.065048e-02  
## hs\_mn\_m\_log2 -7.246265e-03  
## hs\_mo\_c\_log2 -1.097256e-02  
## hs\_mo\_m\_log2 3.141166e-02  
## hs\_pb\_c\_log2 -2.789065e-02  
## hs\_pb\_m\_log2 -7.843584e-03  
## hs\_tl\_cdich\_noneUndetected 4.078968e-02  
## hs\_tl\_mdich\_noneUndetected -2.019338e-01  
## hs\_dde\_cadj\_log2 -1.123795e-02  
## hs\_dde\_madj\_log2 -1.452270e-02  
## hs\_ddt\_cadj\_log2 -1.870973e-03  
## hs\_ddt\_madj\_log2 -1.636478e-03  
## hs\_hcb\_cadj\_log2 -2.161432e-02  
## hs\_hcb\_madj\_log2 -1.648839e-02  
## hs\_pcb118\_cadj\_log2 -1.015967e-02  
## hs\_pcb118\_madj\_log2 1.139057e-02  
## hs\_pcb138\_cadj\_log2 -8.535849e-03  
## hs\_pcb138\_madj\_log2 2.183121e-03  
## hs\_pcb153\_cadj\_log2 -1.379857e-02  
## hs\_pcb153\_madj\_log2 9.355895e-03  
## hs\_pcb170\_cadj\_log2 4.356429e-03  
## hs\_pcb170\_madj\_log2 5.319730e-03  
## hs\_pcb180\_cadj\_log2 -5.681209e-03  
## hs\_pcb180\_madj\_log2 -6.115394e-03  
## hs\_sum\_pc\_bs5\_cadj\_log2 -1.474560e-02  
## hs\_sum\_pc\_bs5\_madj\_log2 7.004742e-03  
## hs\_dep\_cadj\_log2 -3.604075e-03  
## hs\_dep\_madj\_log2 -1.034045e-02  
## hs\_detp\_cadj\_log2 -3.088560e-04  
## hs\_detp\_madj\_log2 -6.346230e-03  
## hs\_dmdtp\_cdich\_noneUndetected 1.244049e-02  
## hs\_dmp\_cadj\_log2 -1.597894e-03  
## hs\_dmp\_madj\_log2 1.983824e-03  
## hs\_dmtp\_cadj\_log2 1.044356e-02  
## hs\_dmtp\_madj\_log2 -3.931627e-03  
## hs\_pbde153\_cadj\_log2 -5.855717e-03  
## hs\_pbde153\_madj\_log2 -1.376512e-03  
## hs\_pbde47\_cadj\_log2 -2.395586e-03  
## hs\_pbde47\_madj\_log2 -1.271719e-02  
## hs\_pfhxs\_c\_log2 8.712197e-03  
## hs\_pfhxs\_m\_log2 4.976116e-03  
## hs\_pfna\_c\_log2 -1.704903e-03  
## hs\_pfna\_m\_log2 -1.682990e-02  
## hs\_pfoa\_c\_log2 1.329143e-02  
## hs\_pfoa\_m\_log2 2.000683e-03  
## hs\_pfos\_c\_log2 1.353752e-02  
## hs\_pfos\_m\_log2 8.119343e-03  
## hs\_pfunda\_c\_log2 9.585448e-03  
## hs\_pfunda\_m\_log2 -1.742107e-02  
## hs\_bpa\_cadj\_log2 -8.735124e-03  
## hs\_bpa\_madj\_log2 -6.300711e-03  
## hs\_bupa\_cadj\_log2 -1.259896e-02  
## hs\_bupa\_madj\_log2 -1.935695e-03  
## hs\_etpa\_cadj\_log2 -3.701026e-03  
## hs\_etpa\_madj\_log2 -7.721548e-03  
## hs\_mepa\_cadj\_log2 2.312457e-03  
## hs\_mepa\_madj\_log2 -6.254681e-04  
## hs\_oxbe\_cadj\_log2 5.210384e-03  
## hs\_oxbe\_madj\_log2 1.708636e-03  
## hs\_prpa\_cadj\_log2 -1.030414e-03  
## hs\_prpa\_madj\_log2 -4.301150e-03  
## hs\_trcs\_cadj\_log2 2.266701e-03  
## hs\_trcs\_madj\_log2 3.667576e-03  
## hs\_mbzp\_cadj\_log2 7.073325e-04  
## hs\_mbzp\_madj\_log2 1.814545e-03  
## hs\_mecpp\_cadj\_log2 -1.996216e-02  
## hs\_mecpp\_madj\_log2 -1.073304e-02  
## hs\_mehhp\_cadj\_log2 8.110137e-04  
## hs\_mehhp\_madj\_log2 -7.386665e-03  
## hs\_mehp\_cadj\_log2 -6.652954e-03  
## hs\_mehp\_madj\_log2 -5.341641e-03  
## hs\_meohp\_cadj\_log2 -7.049916e-03  
## hs\_meohp\_madj\_log2 -1.565561e-02  
## hs\_mep\_cadj\_log2 1.760823e-02  
## hs\_mep\_madj\_log2 -1.195129e-02  
## hs\_mibp\_cadj\_log2 1.943174e-02  
## hs\_mibp\_madj\_log2 2.560121e-03  
## hs\_mnbp\_cadj\_log2 1.304013e-03  
## hs\_mnbp\_madj\_log2 -1.069132e-02  
## hs\_ohminp\_cadj\_log2 1.308471e-02  
## hs\_ohminp\_madj\_log2 2.947996e-04  
## hs\_oxominp\_cadj\_log2 4.699225e-03  
## hs\_oxominp\_madj\_log2 -6.740730e-03  
## hs\_sum\_dehp\_cadj\_log2 -8.151724e-04  
## hs\_sum\_dehp\_madj\_log2 -1.095735e-02  
## fas\_cat\_noneMiddle -4.333681e-02  
## fas\_cat\_noneHigh 6.483059e-03  
## hs\_contactfam\_3cat\_num\_noneOnce a week -3.931092e-03  
## hs\_contactfam\_3cat\_num\_noneLess than once a week 1.914454e-03  
## hs\_hm\_pers\_none 6.973738e-03  
## hs\_participation\_3cat\_none1 organisation 6.536000e-02  
## hs\_participation\_3cat\_none2 or more organisations -2.596306e-02  
## e3\_asmokcigd\_p\_none 3.082957e-03  
## hs\_cotinine\_cdich\_noneUndetected -4.063640e-02  
## hs\_cotinine\_mcat\_noneSHS smokers -1.527645e-02  
## hs\_cotinine\_mcat\_noneSmokers -1.973741e-02  
## hs\_globalexp2\_noneno exposure 1.086391e-02  
## hs\_smk\_parents\_noneneither -4.076928e-02  
## hs\_smk\_parents\_noneone 4.082502e-02  
## h\_bro\_preg\_log 9.968831e-03  
## h\_clf\_preg\_log -2.061128e-03  
## h\_thm\_preg\_log 1.731940e-02  
## e3\_bw 2.594204e-05  
## h\_mbmi\_none 2.100140e-03  
## hs\_wgtgain\_none 7.916860e-04  
## h\_edumc\_none2 -3.380818e-02  
## h\_edumc\_none3 3.147437e-02  
## h\_native\_none1 8.904575e-02  
## h\_native\_none2 -2.582384e-02

varImp(lasso.m)

## glmnet variable importance  
##   
## only 20 most important variables shown (out of 211)  
##   
## Overall  
## hs\_tl\_mdich\_noneUndetected 100.00  
## hs\_pm25abs\_yr\_hs\_h\_log 65.99  
## h\_fish\_preg\_ter(1.9,4.1] 44.28  
## h\_native\_none1 44.09  
## h\_benzene\_log 40.44  
## h\_abs\_ratio\_preg\_log 37.61  
## hs\_participation\_3cat\_none1 organisation 32.36  
## hs\_pet\_cat\_r2\_none1 29.66  
## h\_fastfood\_preg\_ter(0.25,0.83] 29.49  
## h\_fish\_preg\_ter(4.1,Inf] 28.10  
## hs\_fastfood\_ter(0.132,0.5] 26.15  
## h\_pamod\_t3\_noneSometimes 25.32  
## hs\_pm25abs\_wk\_hs\_h\_log 23.22  
## hs\_cu\_m\_log2 22.52  
## h\_fastfood\_preg\_ter(0.83,Inf] 22.26  
## hs\_total\_bread\_ter(7,17.5] 22.11  
## fas\_cat\_noneMiddle 21.45  
## h\_no2\_ratio\_preg\_log 21.25  
## hs\_total\_bread\_ter(17.5,Inf] 20.61  
## hs\_smk\_parents\_noneone 20.21

# Make predictions  
pred.lasso = predict(lasso.m, training)  
pred.lasso.prob = predict(lasso.m, training, type = "prob")  
  
# Model prediction performance  
eval.results = confusionMatrix(pred.lasso, training$hs\_asthma, positive = "1")  
print(eval.results)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 485 20  
## 1 327 80  
##   
## Accuracy : 0.6195   
## 95% CI : (0.5871, 0.6511)  
## No Information Rate : 0.8904   
## P-Value [Acc > NIR] : 1   
##   
## Kappa : 0.1694   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.80000   
## Specificity : 0.59729   
## Pos Pred Value : 0.19656   
## Neg Pred Value : 0.96040   
## Prevalence : 0.10965   
## Detection Rate : 0.08772   
## Detection Prevalence : 0.44627   
## Balanced Accuracy : 0.69865   
##   
## 'Positive' Class : 1   
##

#Accuracy of this model is 0.62

## Final accuracy testing

set.seed(100)  
  
# Using best fit model from above with testing data  
pred.lasso.f = predict(lasso.m, testing)  
pred.lasso.f.prob = predict(lasso.m, testing, type = "prob")  
  
# Evaluating in testing data with confusion matrix  
eval.results = confusionMatrix(pred.lasso.f, testing$hs\_asthma, positive = "1")  
print(eval.results)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 198 17  
## 1 149 25  
##   
## Accuracy : 0.5733   
## 95% CI : (0.5224, 0.623)  
## No Information Rate : 0.892   
## P-Value [Acc > NIR] : 1   
##   
## Kappa : 0.0696   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.59524   
## Specificity : 0.57061   
## Pos Pred Value : 0.14368   
## Neg Pred Value : 0.92093   
## Prevalence : 0.10797   
## Detection Rate : 0.06427   
## Detection Prevalence : 0.44730   
## Balanced Accuracy : 0.58292   
##   
## 'Positive' Class : 1   
##

Final accuracy testing shows that the accuracy of this model is 0.59.

Risk factors show that undetectable thallium levels in the mother is the most important predictor, which tells me that something about this model isn’t quite right. Other important predictors of asthma show that PM25 absorbance the year before the examination, fruit and vegetable intake, and PM absorbance.