DATA 621 Final Project

Farhana Zahir, Vijaya Cherukuri, Scott Reed, Shovon Biswas, Habib Khan, Alain Kuiete Tchoupou

11/18/2020

## OVERVIEW

The self-management of asthma help improve patient health. Asthma self-management provide to the patient and caregivers the skills to understand the disease and its treatment. It teaches them to take medications appropriately, recognize early signs and symptoms of asthma episodes, seek medical care as appropriate, and identify and avoid environmental asthma allergens and irritants In this project, we study the characteristics that influence asthma self-management.

## [1] 13922 899

The data set come from CDC with ulr = “<https://www.cdc.gov/brfss/acbs/2016_documentation.html>”. It is a survey study. The download file is “2016 ACBS Adult Data SAS [ZIP – 3.10 MB]” The unzip file has 899 variables and 13,922 cases. We have selected the variables to use on our studies.

## EXPLORATORY DATA ANALYSIS

Meaning of variables used in the dataset

ASTHNOW Have you ever been told by a doctor or other health professional that you have asthma?

TCH\_SIGN Has a doctor or other health professional ever taught you… a. How to recognize early signs or symptoms of an asthma episode?

TCH\_RESP Has a doctor or other health professional ever taught you… b. What to do during an asthma episode or attack?

TCH\_MON A peak flow meter is a hand held device that measures how quickly you can blow air out of your lungs. Has a doctor or other health professional ever taught you… c. How to use a peak flow meter to adjust your daily medications?

MGT\_PLAN An asthma action plan, or asthma management plan, is a form with instructions about when to change the amount or type of medicine, when to call the doctorfor advice, and when to go to the emergency room. Has a doctor or other health professional EVER given you an asthma action plan?

MOD\_ENV (7.13) INTERVIEWER READ: Now, back to questions specifically about you. Has a health professional ever advised you to change things in your home, school, or work to improve your asthma

MGT\_CLAS Have you ever taken a course or class on how to manage your asthma?

INHALERH (8.3) Did a doctor or other health professional show you how to use the inhaler?

INHALERW (8.4) Did a doctor or other health professional watch you use the inhaler?

Responses types (1) YES (2) NO (7) DON’T KNOW (9) REFUSED

Predictors

MISS\_DAY = “NUMBER OF MISSED DAYS”

MOD\_ENV = “EVER ADVISED CHANGE THINGS IN YOUR HOME”

AGEDX = “AGE AT ASTHMA DIAGNOSIS”

AGEG\_F6\_M = “MODIFIED SIX AGE GROUPS USED IN ASTHMA ADULT POST-STRATIFICATION”

AIRCLEANER = “AIR CLEANER USED”

ASMDCOST = “COST BARRIER: PRIMARY CARE DOCTOR”

ASRXCOST = “COST BARRIER: MEDICATION”

ASSPCOST = “COST BARRIER: SPECIALIST”

CATTMPTS\_F = “DISPOSITION CODES FOR CALL ATTEMPTS 1 THROUGH 20 …”

EMP\_STAT = “CURRENT EMPLOYMENT STATUS”

EPIS\_12M = “ASTHMA EPISODE OR ATTACK”

EPIS\_TP = “NUMBER OF EPISODES / ATTACKS”

ER\_TIMES = “NUMBER OF EMERGENCY ROOM VISITS”

ER\_VISIT = “EMERGENCY ROOM VISIT”

EVER\_ASTH = “EVER HAVE ASTHMA INCONSISTENT WITH BRFSS”

HOSPPLAN = “HOSPITAL FOLLOW-UP”

HOSPTIME = “NUMBER OF HOSPITAL VISITS”

HOSP\_VST = “HOSPITAL VISIT”

QSTLANG\_F = “LANGUAGE IDENTIFIER”

SCR\_MED3 = “HAVE ALL THE MEDICATIONS”

UNEMP\_R = “REASON NOT NOW EMPLOYED”

URG\_TIME = “NUMBER OF URGENT VISITS”

WORKENV5 = “ASTHMA AGGRAVATED BY CURRENT JOB”

WORKENV6 = “ASTHMA CAUSED BY CURRENT JOB”

WORKENV7 = “ASTHMA AGGRAVATED BY PREVIOUS JOB”

WORKENV8 = “ASTHMA CAUSED BY PREVIOUS JOB”

WORKQUIT1 = “EVER CHANGE OR QUIT A JOB”

WORKSEN3 = “DOCTOR DIAGNOSED WORK ASTHMA”

WORKSEN4 = “SELF-IDENTIFIED WORK ASTHMA”

WORKTALK = “DOCTOR DISCUSSED WORK ASTHMA”

INS1 = “INSURANCE”

INS2 = “INSURANCE OR COVERAGE GAP”

LASTSYMP = “LAST HAD ANY SYMPTOMS OF ASTHMA”

LAST\_MD = “LAST TALKED TO A DOCTOR”

LAST\_MED = “LAST TOOK ASTHMA MEDICATION”

COMPASTH = “TYPICAL ATTACK”

### Constructing the Data Frame by Selecting variables

We select all possible variable that we can use in our dataset. We also start to clean the dataset

## TCH.SIGN TCH.RESP TCH.MON MGT.PLAN MGT.CLAS INHALERW MOD.ENV SEX   
## 1:8639 1:9936 1:5694 1:3814 1: 1190 1:9382 1:4401 1:4786   
## 2:4972 2:3669 2:8014 2:9759 2:12675 2:2925 2:9407 2:9136   
## 7: 301 7: 299 7: 195 7: 336 7: 53 5: 386 7: 110   
## 9: 10 9: 18 9: 19 9: 13 9: 4 6: 762 9: 4   
## 7: 466   
## 9: 1   
##   
## AGEG.F7 X\_RACEGR3 EDUCAL X\_INCOMG X\_RFBMI5 SMOKE100 COPD   
## 1: 763 1:10806 1: 24 1:1910 1:3640 1:6546 1 : 2686   
## 2:1186 2: 766 2: 350 2:2264 2:9647 2:7312 2 :11027   
## 3:1419 3: 513 3: 722 3:1272 9: 635 7: 59 7 : 149   
## 4:2149 4: 459 4:3307 4:1659 9: 5 9 : 16   
## 5:3476 5: 1209 5:4177 5:5265 NA's: 44   
## 6:3199 9: 169 6:5321 9:1552   
## 7:1730 9: 21   
## EMPHY DEPRESS BRONCH SYMP.30D DUR.30D INCINDT   
## 1 : 1133 1 :5185 1 : 3670 66 :4351 12 :4735 1: 316   
## 2 :12638 2 :8628 2 :10027 30 :2237 6 :4351 2: 1002   
## 7 : 93 7 : 36 7 : 166 100 :1618 10 :1618 3:12549   
## 9 : 14 9 : 29 9 : 15 88 : 749 1 :1288 7: 44   
## NA's: 44 NA's: 44 NA's: 44 2 : 547 2 : 899 9: 11   
## 3 : 512 11 : 749   
## (Other):3908 (Other): 282   
## LAST.MD LAST.MED LAST.SYMP EPIS.12M COMPASTH INS1   
## 4 :7977 1 :4924 1 :3567 1:5210 11 :4361 1:13121   
## 5 :1820 7 :2978 7 :2616 2:4237 6 :4351 2: 767   
## 6 : 819 3 :1555 3 :2515 6:4351 3 :3205 7: 26   
## 7 :3025 4 :1238 4 :1618 7: 117 1 :1169 9: 8   
## 77: 131 5 :1072 2 :1613 9: 7 2 : 781   
## 88: 135 2 :1059 5 :1113 7 : 41   
## 99: 15 (Other):1096 (Other): 880 (Other): 14   
## INS2 ER.VISIT HOSP.VST ASMDCOST ASRXCOST ASSPCOST WORKTALK   
## 1: 683 1:1347 1: 380 1 : 770 1 :1559 1 : 506 1 : 2592   
## 2:12415 2:6700 2:6702 2 :10370 2 :9592 2 :10647 2 :10851   
## 5: 767 5:2737 4: 979 5 : 2736 5 :2736 5 : 2736 6 : 281   
## 7: 46 6:3105 5:2737 7 : 31 7 : 18 7 : 15 7 : 144   
## 9: 11 7: 32 6:3105 9 : 10 9 : 12 9 : 13 8 : 31   
## 9: 1 7: 19 NA's: 5 NA's: 5 NA's: 5 9 : 11   
## NA's: 12

## [1] 11494 34

## Structure of the data

## 'data.frame': 11494 obs. of 34 variables:  
## $ TCH.SIGN : num 1 2 1 2 2 1 1 2 1 2 ...  
## ..- attr(\*, "label")= chr "EVER TAUGHT RECOGNIZE EARLY SIGN OR SYMPTOMS"  
## ..- attr(\*, "format.sas")= chr "TCH\_SIGN"  
## $ TCH.RESP : num 1 1 1 2 1 1 1 1 1 1 ...  
## ..- attr(\*, "label")= chr "EVER TAUGHT WHAT TO DO DURING ASTHMA EPISODE OR ATTACK"  
## ..- attr(\*, "format.sas")= chr "TCH\_RESP"  
## $ TCH.MON : num 2 2 2 2 2 1 1 2 2 2 ...  
## ..- attr(\*, "label")= chr "EVER TAUGHT HOW TO USE A PEAK FLOW"  
## ..- attr(\*, "format.sas")= chr "TCH\_MON"  
## $ MGT.PLAN : num 2 2 2 2 2 2 1 2 2 2 ...  
## ..- attr(\*, "label")= chr "EVER GIVEN AN ASTHMA ACTION PLAN"  
## ..- attr(\*, "format.sas")= chr "MGT\_PLAN"  
## $ MGT.CLAS : num 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "EVER TAKEN A COURSE TO MANAGE ASTHMA"  
## ..- attr(\*, "format.sas")= chr "MGT\_CLAS"  
## $ INHALERW : num 2 2 1 1 1 1 1 1 1 1 ...  
## ..- attr(\*, "label")= chr "INHALER USE WATCHED"  
## ..- attr(\*, "format.sas")= chr "INHALERW"  
## $ MOD.ENV : num 2 2 2 2 1 2 2 2 1 2 ...  
## ..- attr(\*, "label")= chr "EVER ADVISED CHANGE THINGS IN YOUR HOME"  
## ..- attr(\*, "format.sas")= chr "MOD\_ENV"  
## $ SEX : num 1 2 2 2 2 2 1 2 2 2 ...  
## ..- attr(\*, "label")= chr "RESPONDENTS SEX"  
## ..- attr(\*, "format.sas")= chr "SEX"  
## $ AGEG.F7 : num 4 5 5 3 6 5 4 6 6 7 ...  
## ..- attr(\*, "label")= chr "AGE COLLAPSED TO 7 GROUPS FOR ASTHMA CALL-BACK"  
## ..- attr(\*, "format.sas")= chr "AGEG\_F7Z"  
## $ X\_RACEGR3: num 3 1 1 5 1 5 1 1 1 1 ...  
## ..- attr(\*, "label")= chr "COMPUTED FIVE LEVEL RACE/ETHNICITY CATEGORY."  
## ..- attr(\*, "format.sas")= chr "\_3RACEGR"  
## $ EDUCAL : num 6 4 4 5 6 6 6 6 6 5 ...  
## ..- attr(\*, "label")= chr "EDUCATION LEVEL"  
## ..- attr(\*, "format.sas")= chr "EDUCA"  
## $ X\_INCOMG : num 5 1 1 5 5 5 5 5 3 9 ...  
## ..- attr(\*, "label")= chr "COMPUTED INCOME CATEGORIES"  
## ..- attr(\*, "format.sas")= chr "\_INCOMG"  
## $ X\_RFBMI5 : num 2 2 2 2 2 2 1 2 2 1 ...  
## ..- attr(\*, "label")= chr "OVERWEIGHT OR OBESE CALCULATED VARIABLE"  
## ..- attr(\*, "format.sas")= chr "\_5RFBMI"  
## $ SMOKE100 : num 2 1 1 2 1 2 1 1 2 2 ...  
## ..- attr(\*, "label")= chr "SMOKED AT LEAST 100 CIGARETTES"  
## ..- attr(\*, "format.sas")= chr "SMOK100\_"  
## $ COPD : num 2 1 2 2 2 2 2 2 2 1 ...  
## ..- attr(\*, "label")= chr "EVER TOLD HAVE CHRONIC OBSTRUCTIVE PULMONARY DISEASE"  
## ..- attr(\*, "format.sas")= chr "COPD"  
## $ EMPHY : num 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "EVER TOLD HAVE EMPHYSEMA"  
## ..- attr(\*, "format.sas")= chr "EMPHY"  
## $ DEPRESS : num 2 1 2 2 2 2 2 2 1 1 ...  
## ..- attr(\*, "label")= chr "EVER TOLD DEPRESSED"  
## ..- attr(\*, "format.sas")= chr "DEPRESS"  
## $ BRONCH : num 2 1 2 2 1 2 2 2 1 2 ...  
## ..- attr(\*, "label")= chr "EVER TOLD HAVE CHRONIC BRONCHITIS"  
## ..- attr(\*, "format.sas")= chr "BRONCH"  
## $ SYMP.30D : num 100 30 7 66 17 100 10 66 30 66 ...  
## ..- attr(\*, "label")= chr "SYMPTOM DAYS"  
## ..- attr(\*, "format.sas")= chr "SYMP\_30D"  
## $ DUR.30D : num 10 2 12 6 12 10 12 6 1 6 ...  
## ..- attr(\*, "label")= chr "CONSTANT SYMPTOMS"  
## ..- attr(\*, "format.sas")= chr "DUR\_30D"  
## $ INCINDT : num 3 2 3 3 3 2 3 3 3 3 ...  
## ..- attr(\*, "label")= chr "TIME SINCE DIAGNOSIS"  
## ..- attr(\*, "format.sas")= chr "INCIDNT"  
## $ LAST.MD : num 5 4 4 7 4 4 4 5 4 5 ...  
## ..- attr(\*, "label")= chr "LAST TALKED TO A DOCTOR"  
## ..- attr(\*, "format.sas")= chr "LAST\_MD"  
## $ LAST.MED : num 4 1 3 7 3 1 1 6 1 5 ...  
## ..- attr(\*, "label")= chr "LAST TOOK ASTHMA MEDICATION"  
## ..- attr(\*, "format.sas")= chr "LAST\_MED"  
## $ LAST.SYMP: num 4 1 3 7 3 4 3 5 1 5 ...  
## ..- attr(\*, "label")= chr "LAST HAD ANY SYMPTOMS OF ASTHMA"  
## ..- attr(\*, "format.sas")= chr "LASTSYMP"  
## $ EPIS.12M : num 1 1 1 6 1 2 1 6 2 6 ...  
## ..- attr(\*, "label")= chr "ASTHMA EPISODE OR ATTACK"  
## ..- attr(\*, "format.sas")= chr "EPIS\_12M"  
## $ COMPASTH : num 1 3 1 6 3 11 3 6 11 6 ...  
## ..- attr(\*, "label")= chr "TYPICAL ATTACK"  
## ..- attr(\*, "format.sas")= chr "COMPASTH"  
## $ INS1 : num 1 1 1 2 1 1 1 1 1 1 ...  
## ..- attr(\*, "label")= chr "INSURANCE"  
## ..- attr(\*, "format.sas")= chr "INS1Z"  
## $ INS2 : num 2 2 2 5 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "INSURANCE OR COVERAGE GAP"  
## ..- attr(\*, "format.sas")= chr "INS2Z"  
## $ ER.VISIT : num 6 2 2 5 2 2 2 5 2 6 ...  
## ..- attr(\*, "label")= chr "EMERGENCY ROOM VISIT"  
## ..- attr(\*, "format.sas")= chr "ER\_VISIT"  
## $ HOSP.VST : num 6 2 2 5 2 2 2 5 2 6 ...  
## ..- attr(\*, "label")= chr "HOSPITAL VISIT"  
## ..- attr(\*, "format.sas")= chr "HOSP\_VST"  
## $ ASMDCOST : num 2 2 2 5 2 2 2 5 2 2 ...  
## ..- attr(\*, "label")= chr "COST BARRIER: PRIMARY CARE DOCTOR"  
## ..- attr(\*, "format.sas")= chr "ASMDCOST"  
## $ ASRXCOST : num 2 2 2 5 2 2 2 5 1 2 ...  
## ..- attr(\*, "label")= chr "COST BARRIER: MEDICATION"  
## ..- attr(\*, "format.sas")= chr "ASRXCOST"  
## $ ASSPCOST : num 2 2 2 5 2 2 2 5 2 2 ...  
## ..- attr(\*, "label")= chr "COST BARRIER: SPECIALIST"  
## ..- attr(\*, "format.sas")= chr "ASSPCOST"  
## $ WORKTALK : num 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "DOCTOR DISCUSSED WORK ASTHMA"  
## ..- attr(\*, "format.sas")= chr "WORKTALK"

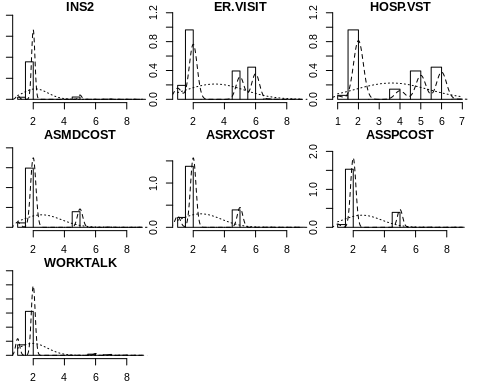
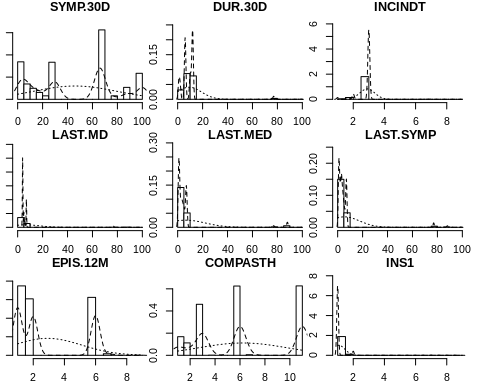
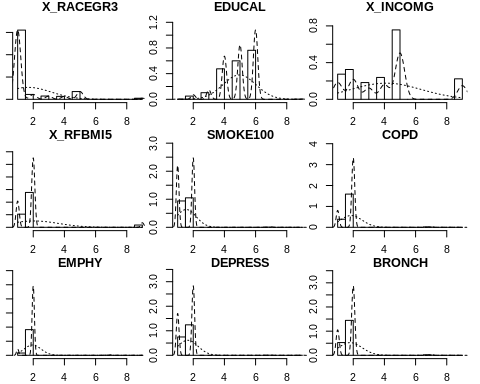
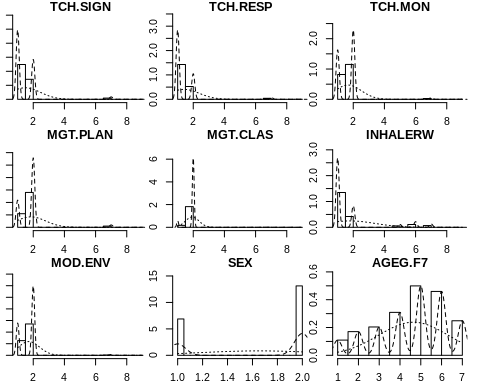
### Summary of the Data

## TCH.SIGN TCH.RESP TCH.MON MGT.PLAN   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000   
## Median :1.000 Median :1.000 Median :2.000 Median :2.000   
## Mean :1.337 Mean :1.238 Mean :1.555 Mean :1.694   
## 3rd Qu.:2.000 3rd Qu.:1.000 3rd Qu.:2.000 3rd Qu.:2.000   
## Max. :2.000 Max. :2.000 Max. :2.000 Max. :2.000   
##   
## MGT.CLAS INHALERW MOD.ENV SEX AGEG.F7   
## Min. :1.000 Min. :1.00 Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:2.000 1st Qu.:1.00 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:4.000   
## Median :2.000 Median :1.00 Median :2.000 Median :2.000 Median :5.000   
## Mean :1.906 Mean :1.24 Mean :1.662 Mean :1.678 Mean :4.594   
## 3rd Qu.:2.000 3rd Qu.:1.00 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:6.000   
## Max. :2.000 Max. :2.00 Max. :2.000 Max. :2.000 Max. :7.000   
##   
## X\_RACEGR3 EDUCAL X\_INCOMG X\_RFBMI5 SMOKE100   
## Min. :1.000 Min. :1.00 Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:4.00 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:1.000   
## Median :1.000 Median :5.00 Median :4.000 Median :2.000 Median :2.000   
## Mean :1.653 Mean :4.98 Mean :4.057 Mean :2.062 Mean :1.549   
## 3rd Qu.:1.000 3rd Qu.:6.00 3rd Qu.:5.000 3rd Qu.:2.000 3rd Qu.:2.000   
## Max. :9.000 Max. :9.00 Max. :9.000 Max. :9.000 Max. :9.000   
##   
## COPD EMPHY DEPRESS BRONCH   
## Min. :1.00 Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:2.00 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:1.000   
## Median :2.00 Median :2.000 Median :2.000 Median :2.000   
## Mean :1.85 Mean :1.945 Mean :1.633 Mean :1.769   
## 3rd Qu.:2.00 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000   
## Max. :9.00 Max. :9.000 Max. :9.000 Max. :9.000   
## NA's :30 NA's :30 NA's :30 NA's :30   
## SYMP.30D DUR.30D INCINDT LAST.MD   
## Min. : 1.00 Min. : 1.000 Min. :1.000 Min. : 4.000   
## 1st Qu.: 10.00 1st Qu.: 6.000 1st Qu.:3.000 1st Qu.: 4.000   
## Median : 30.00 Median :10.000 Median :3.000 Median : 4.000   
## Mean : 44.46 Mean : 9.201 Mean :2.892 Mean : 5.729   
## 3rd Qu.: 66.00 3rd Qu.:12.000 3rd Qu.:3.000 3rd Qu.: 5.000   
## Max. :100.00 Max. :99.000 Max. :9.000 Max. :99.000   
##   
## LAST.MED LAST.SYMP EPIS.12M COMPASTH   
## Min. : 1.000 Min. : 1.000 Min. :1.000 Min. : 1.000   
## 1st Qu.: 1.000 1st Qu.: 1.000 1st Qu.:1.000 1st Qu.: 3.000   
## Median : 3.000 Median : 3.000 Median :2.000 Median : 6.000   
## Mean : 3.733 Mean : 4.742 Mean :2.703 Mean : 6.126   
## 3rd Qu.: 5.000 3rd Qu.: 5.000 3rd Qu.:6.000 3rd Qu.:11.000   
## Max. :99.000 Max. :99.000 Max. :9.000 Max. :11.000   
##   
## INS1 INS2 ER.VISIT HOSP.VST   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000   
## Median :1.000 Median :2.000 Median :2.000 Median :2.000   
## Mean :1.065 Mean :2.127 Mean :3.256 Mean :3.468   
## 3rd Qu.:1.000 3rd Qu.:2.000 3rd Qu.:5.000 3rd Qu.:5.000   
## Max. :9.000 Max. :9.000 Max. :9.000 Max. :7.000   
##   
## ASMDCOST ASRXCOST ASSPCOST WORKTALK   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000   
## Median :2.000 Median :2.000 Median :2.000 Median :2.000   
## Mean :2.416 Mean :2.353 Mean :2.433 Mean :1.928   
## 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000   
## Max. :9.000 Max. :9.000 Max. :9.000 Max. :9.000   
## NA's :4 NA's :4 NA's :4 NA's :8

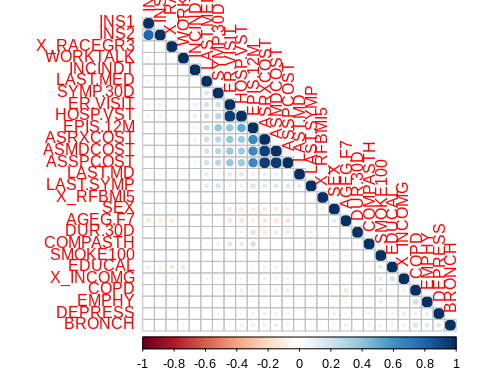
### Distribution of the Variables in the Data

#### Histograms

Histograms tell us how the data is distributed in the dataset (numeric fields).



### The correlations betweeen predictors



There are highly correlated predictors. We are going to remove some of them.

## [1] "TCH.SIGN" "TCH.RESP" "TCH.MON" "MGT.PLAN" "MGT.CLAS" "INHALERW"   
## [7] "MOD.ENV" "SEX" "AGEG.F7" "X\_RACEGR3" "EDUCAL" "X\_INCOMG"   
## [13] "X\_RFBMI5" "SMOKE100" "COPD" "EMPHY" "DEPRESS" "BRONCH"   
## [19] "SYMP.30D" "DUR.30D" "INCINDT" "LAST.MD" "LAST.MED" "LAST.SYMP"  
## [25] "EPIS.12M" "COMPASTH" "INS1" "INS2" "ER.VISIT" "HOSP.VST"   
## [31] "ASRXCOST" "WORKTALK"

### CONSTRUCT THE RESPONSE VARIABLE

We first extract variables related to education, #### Selection of variables

## TCH.SIGN TCH.RESP TCH.MON MGT.PLAN MGT.CLAS INHALERW MOD.ENV  
## 1 1 1 2 2 2 2 2  
## 2 2 1 2 2 2 2 2  
## 3 1 1 2 2 2 1 2  
## 4 2 2 2 2 2 1 2  
## 5 2 1 2 2 2 1 1  
## 6 1 1 1 2 2 1 2

#### Exploration of the clustering

## TCH.SIGN TCH.RESP TCH.MON MGT.PLAN MGT.CLAS INHALERW MOD.ENV   
## 1:7621 1:8760 1:5118 1:3522 1: 1078 1:8738 1:3889   
## 2:3873 2:2734 2:6376 2:7972 2:10416 2:2756 2:7605

#### Elbow method to find the number of clusters

We run kmeans with different clusters from 1 to 16 and we produce a scree plot to determine the number of cluster at the elbow. 

The number of cluster is 3

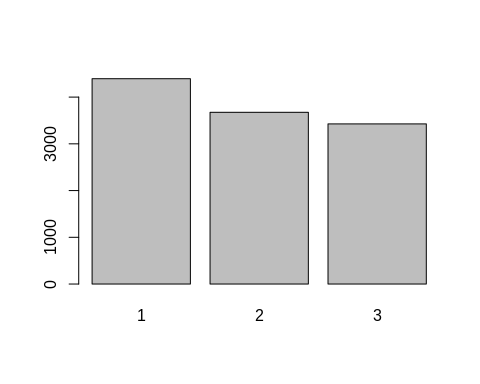
#### Now we do the clustering and extract the centers of resulting model

## TCH.SIGN TCH.RESP TCH.MON MGT.PLAN MGT.CLAS INHALERW MOD.ENV  
## 1 1.057819 1.034145 1.000000 1.427271 1.828819 1.108127 1.532666  
## 2 1.977681 1.648612 1.802667 1.950463 1.979314 1.429777 1.822265  
## 3 1.007879 1.058652 2.000000 1.759556 1.927050 1.204844 1.654800

#### We add the point classification to the original data

## TCH.SIGN TCH.RESP TCH.MON MGT.PLAN MGT.CLAS INHALERW MOD.ENV target  
## 1 1 1 2 2 2 2 2 3  
## 2 2 1 2 2 2 2 2 2  
## 3 1 1 2 2 2 1 2 3  
## 4 2 2 2 2 2 1 2 2  
## 5 2 1 2 2 2 1 1 2  
## 6 1 1 1 2 2 1 2 1

## TCH.SIGN TCH.RESP TCH.MON MGT.PLAN MGT.CLAS INHALERW MOD.ENV target   
## 1:7621 1:8760 1:5118 1:3522 1: 1078 1:8738 1:3889 1:4393   
## 2:3873 2:2734 2:6376 2:7972 2:10416 2:2756 2:7605 2:3674   
## 3:3427



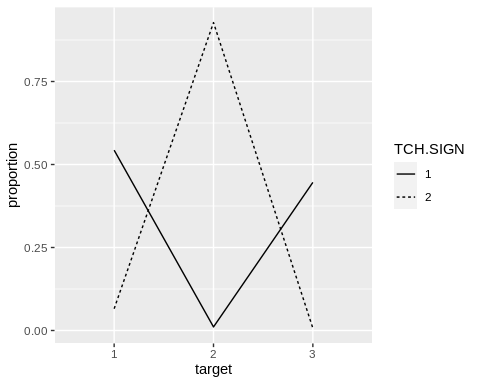
View of the clustering result

### Interpretation of the Selft-Management Response clustering

#### TCH.SIGN

## # A tibble: 6 x 5  
## TCH.SIGN target count etotal proportion  
## <fct> <fct> <int> <int> <dbl>  
## 1 1 1 4139 7621 0.543   
## 2 1 2 82 7621 0.0108   
## 3 1 3 3400 7621 0.446   
## 4 2 1 254 3873 0.0656   
## 5 2 2 3592 3873 0.927   
## 6 2 3 27 3873 0.00697

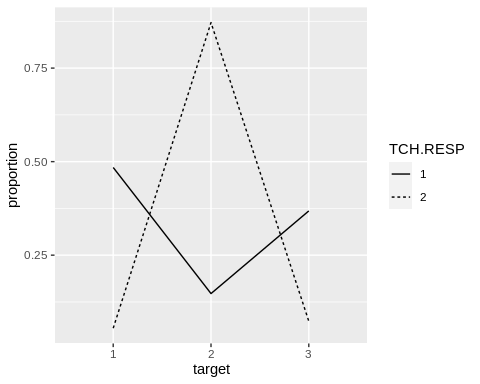
## # A tibble: 2 x 6  
## # Groups: target [2]  
## TCH.SIGN target count etotal proportion group.max  
## <fct> <fct> <int> <int> <dbl> <int>  
## 1 1 1 4139 7621 0.543 4139  
## 2 2 2 3592 3873 0.927 3592

 In the target response, 8 is the positive answer, 3 is the negative answer, 5 is don’t know and 6 is refused for the question: TCH\_SIGN Has a doctor or other health professional ever taught you… a. How to recognize early signs or symptoms of an asthma episode?

#### TCH.RESP

## # A tibble: 6 x 5  
## TCH.RESP target count etotal proportion  
## <fct> <fct> <int> <int> <dbl>  
## 1 1 1 4243 8760 0.484   
## 2 1 2 1291 8760 0.147   
## 3 1 3 3226 8760 0.368   
## 4 2 1 150 2734 0.0549  
## 5 2 2 2383 2734 0.872   
## 6 2 3 201 2734 0.0735

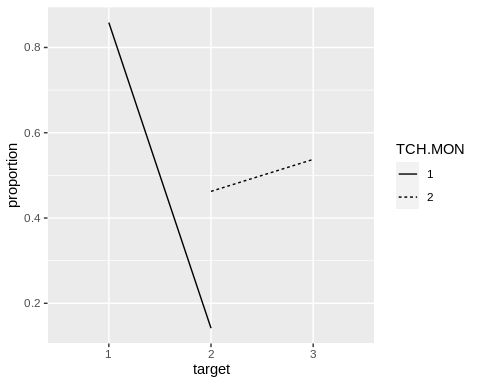
## # A tibble: 2 x 6  
## # Groups: target [2]  
## TCH.RESP target count etotal proportion group.max  
## <fct> <fct> <int> <int> <dbl> <int>  
## 1 1 1 4243 8760 0.484 4243  
## 2 2 2 2383 2734 0.872 2383

 In the target response, 8 is the positive answer, 3 is the negative answer, 1 is don’t know and 1 is refused for the question: TCH\_RESP Has a doctor or other health professional ever taught you… b. What to do during an asthma episode or attack?

#### TCH.MON

## # A tibble: 4 x 5  
## TCH.MON target count etotal proportion  
## <fct> <fct> <int> <int> <dbl>  
## 1 1 1 4393 5118 0.858  
## 2 1 2 725 5118 0.142  
## 3 2 2 2949 6376 0.463  
## 4 2 3 3427 6376 0.537

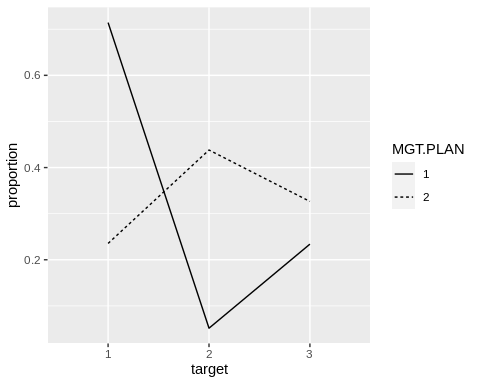
## # A tibble: 2 x 6  
## # Groups: target [2]  
## TCH.MON target count etotal proportion group.max  
## <fct> <fct> <int> <int> <dbl> <int>  
## 1 1 1 4393 5118 0.858 4393  
## 2 2 3 3427 6376 0.537 3427

 In the target response, 8 is the positive answer, 7 are the negative answers, 2 is don’t know and 2 is refused for the question: TCH\_MON A peak flow meter is a hand held device that measures how quickly you can blow air out of your lungs. Has a doctor or other health professional ever taught you… c. How to use a peak flow meter to adjust your daily medications?

#### MGT.PLAN

## # A tibble: 6 x 5  
## MGT.PLAN target count etotal proportion  
## <fct> <fct> <int> <int> <dbl>  
## 1 1 1 2516 3522 0.714   
## 2 1 2 182 3522 0.0517  
## 3 1 3 824 3522 0.234   
## 4 2 1 1877 7972 0.235   
## 5 2 2 3492 7972 0.438   
## 6 2 3 2603 7972 0.327

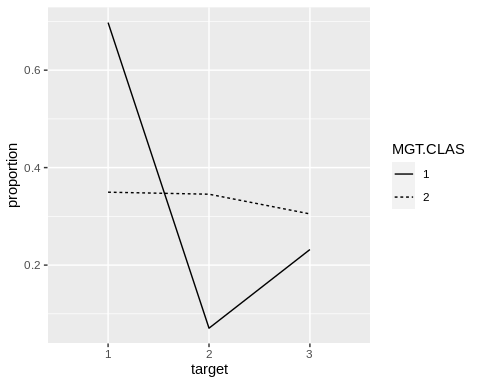
## # A tibble: 2 x 6  
## # Groups: target [2]  
## MGT.PLAN target count etotal proportion group.max  
## <fct> <fct> <int> <int> <dbl> <int>  
## 1 1 1 2516 3522 0.714 2516  
## 2 2 2 3492 7972 0.438 3492

 In the target response, 8 is the positive answer, 3 is the negative answer, 9 is don’t know and 9 is refused for the question: MGT\_PLAN An asthma action plan, or asthma management plan, is a form with instructions about when to change the amount or type of medicine, when to call the doctor for advice, and when to go to the emergency room. Has a doctor or other health professional EVER given you an asthma action plan?

#### MGT.CLAS

## # A tibble: 6 x 5  
## MGT.CLAS target count etotal proportion  
## <fct> <fct> <int> <int> <dbl>  
## 1 1 1 752 1078 0.698   
## 2 1 2 76 1078 0.0705  
## 3 1 3 250 1078 0.232   
## 4 2 1 3641 10416 0.350   
## 5 2 2 3598 10416 0.345   
## 6 2 3 3177 10416 0.305

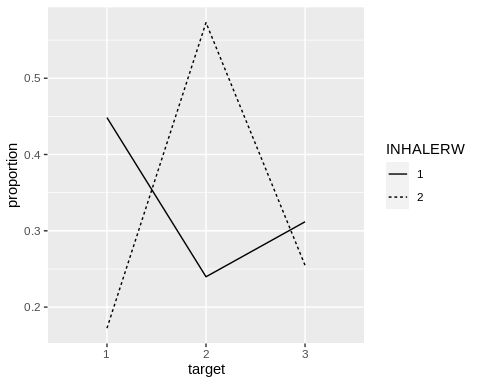
## # A tibble: 2 x 6  
## # Groups: target [1]  
## MGT.CLAS target count etotal proportion group.max  
## <fct> <fct> <int> <int> <dbl> <int>  
## 1 1 1 752 1078 0.698 752  
## 2 2 1 3641 10416 0.350 3641

 In the target response, 8 is the positive answer, 8 or(3,7) is the negative answer, 8 is don’t know and 6 is refused for the question: MGT\_CLAS Have you ever taken a course or class on how to manage your asthma?

#### INHALERW

## # A tibble: 6 x 5  
## INHALERW target count etotal proportion  
## <fct> <fct> <int> <int> <dbl>  
## 1 1 1 3918 8738 0.448  
## 2 1 2 2095 8738 0.240  
## 3 1 3 2725 8738 0.312  
## 4 2 1 475 2756 0.172  
## 5 2 2 1579 2756 0.573  
## 6 2 3 702 2756 0.255

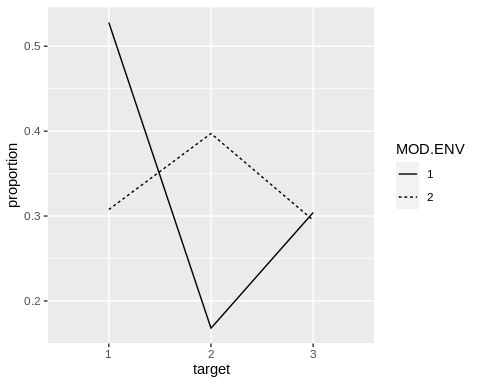
## # A tibble: 2 x 6  
## # Groups: target [2]  
## INHALERW target count etotal proportion group.max  
## <fct> <fct> <int> <int> <dbl> <int>  
## 1 1 1 3918 8738 0.448 3918  
## 2 2 2 1579 2756 0.573 1579

 In the target response, 8 is the positive answer, 3 is the negative answer, 4 is don’t know and 1 is refused for the question: INHALERW (8.4) Did a doctor or other health professional watch you use the inhaler?

#### MOD.ENV

## # A tibble: 6 x 5  
## MOD.ENV target count etotal proportion  
## <fct> <fct> <int> <int> <dbl>  
## 1 1 1 2053 3889 0.528  
## 2 1 2 653 3889 0.168  
## 3 1 3 1183 3889 0.304  
## 4 2 1 2340 7605 0.308  
## 5 2 2 3021 7605 0.397  
## 6 2 3 2244 7605 0.295

## # A tibble: 2 x 6  
## # Groups: target [2]  
## MOD.ENV target count etotal proportion group.max  
## <fct> <fct> <int> <int> <dbl> <int>  
## 1 1 1 2053 3889 0.528 2053  
## 2 2 2 3021 7605 0.397 3021



#### Summary of the response variables

## # A tibble: 2 x 8  
## RESPONSE TCH.SIGN TCH.RES TCH.MON MGT.PLAN MGT.CLAS INHALERW MOD.ENV  
## <chr> <fct> <fct> <fct> <fct> <fct> <fct> <fct>   
## 1 1=YES 1 1 1 1 1 1 1   
## 2 2=NO 2 2 3 2 1 2 2

##### For the response variable TARGET, an excellent management skill has number 2 but a poor management skill has number 1 and 3.

##### We can build a logistics regression on the dataset.

### Here we remove the varibles used to calculate the target variable and reformat the data frame.

## 'data.frame': 11494 obs. of 26 variables:  
## $ TARGET : num 0 1 0 1 1 0 0 1 0 1 ...  
## $ SEX : num 1 2 2 2 2 2 1 2 2 2 ...  
## ..- attr(\*, "label")= chr "RESPONDENTS SEX"  
## ..- attr(\*, "format.sas")= chr "SEX"  
## $ AGEG.F7 : num 4 5 5 3 6 5 4 6 6 7 ...  
## ..- attr(\*, "label")= chr "AGE COLLAPSED TO 7 GROUPS FOR ASTHMA CALL-BACK"  
## ..- attr(\*, "format.sas")= chr "AGEG\_F7Z"  
## $ X\_RACEGR3: num 3 1 1 5 1 5 1 1 1 1 ...  
## ..- attr(\*, "label")= chr "COMPUTED FIVE LEVEL RACE/ETHNICITY CATEGORY."  
## ..- attr(\*, "format.sas")= chr "\_3RACEGR"  
## $ EDUCAL : num 6 4 4 5 6 6 6 6 6 5 ...  
## ..- attr(\*, "label")= chr "EDUCATION LEVEL"  
## ..- attr(\*, "format.sas")= chr "EDUCA"  
## $ X\_INCOMG : num 5 1 1 5 5 5 5 5 3 9 ...  
## ..- attr(\*, "label")= chr "COMPUTED INCOME CATEGORIES"  
## ..- attr(\*, "format.sas")= chr "\_INCOMG"  
## $ X\_RFBMI5 : num 2 2 2 2 2 2 1 2 2 1 ...  
## ..- attr(\*, "label")= chr "OVERWEIGHT OR OBESE CALCULATED VARIABLE"  
## ..- attr(\*, "format.sas")= chr "\_5RFBMI"  
## $ SMOKE100 : num 2 1 1 2 1 2 1 1 2 2 ...  
## ..- attr(\*, "label")= chr "SMOKED AT LEAST 100 CIGARETTES"  
## ..- attr(\*, "format.sas")= chr "SMOK100\_"  
## $ COPD : num 2 1 2 2 2 2 2 2 2 1 ...  
## ..- attr(\*, "label")= chr "EVER TOLD HAVE CHRONIC OBSTRUCTIVE PULMONARY DISEASE"  
## ..- attr(\*, "format.sas")= chr "COPD"  
## $ EMPHY : num 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "EVER TOLD HAVE EMPHYSEMA"  
## ..- attr(\*, "format.sas")= chr "EMPHY"  
## $ DEPRESS : num 2 1 2 2 2 2 2 2 1 1 ...  
## ..- attr(\*, "label")= chr "EVER TOLD DEPRESSED"  
## ..- attr(\*, "format.sas")= chr "DEPRESS"  
## $ BRONCH : num 2 1 2 2 1 2 2 2 1 2 ...  
## ..- attr(\*, "label")= chr "EVER TOLD HAVE CHRONIC BRONCHITIS"  
## ..- attr(\*, "format.sas")= chr "BRONCH"  
## $ SYMP.30D : num 100 30 7 66 17 100 10 66 30 66 ...  
## ..- attr(\*, "label")= chr "SYMPTOM DAYS"  
## ..- attr(\*, "format.sas")= chr "SYMP\_30D"  
## $ DUR.30D : num 10 2 12 6 12 10 12 6 1 6 ...  
## ..- attr(\*, "label")= chr "CONSTANT SYMPTOMS"  
## ..- attr(\*, "format.sas")= chr "DUR\_30D"  
## $ INCINDT : num 3 2 3 3 3 2 3 3 3 3 ...  
## ..- attr(\*, "label")= chr "TIME SINCE DIAGNOSIS"  
## ..- attr(\*, "format.sas")= chr "INCIDNT"  
## $ LAST.MD : num 5 4 4 7 4 4 4 5 4 5 ...  
## ..- attr(\*, "label")= chr "LAST TALKED TO A DOCTOR"  
## ..- attr(\*, "format.sas")= chr "LAST\_MD"  
## $ LAST.MED : num 4 1 3 7 3 1 1 6 1 5 ...  
## ..- attr(\*, "label")= chr "LAST TOOK ASTHMA MEDICATION"  
## ..- attr(\*, "format.sas")= chr "LAST\_MED"  
## $ LAST.SYMP: num 4 1 3 7 3 4 3 5 1 5 ...  
## ..- attr(\*, "label")= chr "LAST HAD ANY SYMPTOMS OF ASTHMA"  
## ..- attr(\*, "format.sas")= chr "LASTSYMP"  
## $ EPIS.12M : num 1 1 1 6 1 2 1 6 2 6 ...  
## ..- attr(\*, "label")= chr "ASTHMA EPISODE OR ATTACK"  
## ..- attr(\*, "format.sas")= chr "EPIS\_12M"  
## $ COMPASTH : num 1 3 1 6 3 11 3 6 11 6 ...  
## ..- attr(\*, "label")= chr "TYPICAL ATTACK"  
## ..- attr(\*, "format.sas")= chr "COMPASTH"  
## $ INS1 : num 1 1 1 2 1 1 1 1 1 1 ...  
## ..- attr(\*, "label")= chr "INSURANCE"  
## ..- attr(\*, "format.sas")= chr "INS1Z"  
## $ INS2 : num 2 2 2 5 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "INSURANCE OR COVERAGE GAP"  
## ..- attr(\*, "format.sas")= chr "INS2Z"  
## $ ER.VISIT : num 6 2 2 5 2 2 2 5 2 6 ...  
## ..- attr(\*, "label")= chr "EMERGENCY ROOM VISIT"  
## ..- attr(\*, "format.sas")= chr "ER\_VISIT"  
## $ HOSP.VST : num 6 2 2 5 2 2 2 5 2 6 ...  
## ..- attr(\*, "label")= chr "HOSPITAL VISIT"  
## ..- attr(\*, "format.sas")= chr "HOSP\_VST"  
## $ ASRXCOST : num 2 2 2 5 2 2 2 5 1 2 ...  
## ..- attr(\*, "label")= chr "COST BARRIER: MEDICATION"  
## ..- attr(\*, "format.sas")= chr "ASRXCOST"  
## $ WORKTALK : num 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "DOCTOR DISCUSSED WORK ASTHMA"  
## ..- attr(\*, "format.sas")= chr "WORKTALK"

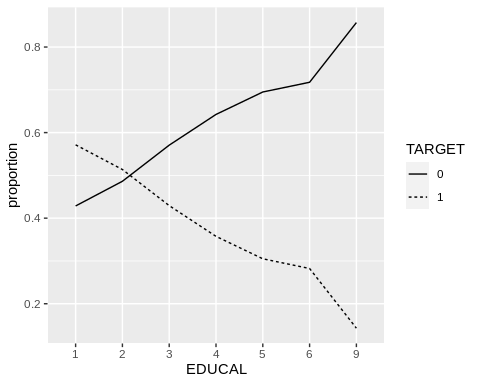
### PREPARE THE DATA FOR MODELISATION

#### We remove the rows with missing values.

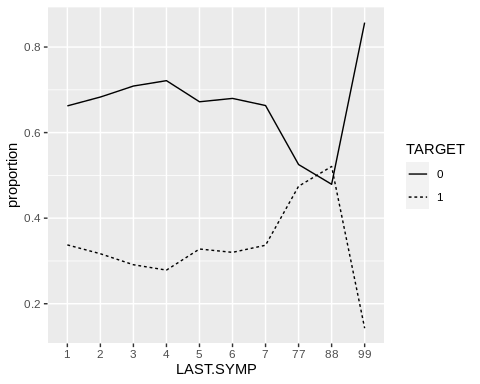
Here were are going to drop missing data because they are only 12 over 13,922 rows. We also transform all predictors to categorical.

## TARGET SEX AGEG.F7 X\_RACEGR3 EDUCAL X\_INCOMG X\_RFBMI5 SMOKE100  
## 0:7799 1:3686 1: 622 1:8935 1: 14 1:1583 1:2917 1:5401   
## 1:3665 2:7778 2: 996 2: 636 2: 255 2:1872 2:8027 2:6017   
## 3:1217 3: 420 3: 573 3:1042 9: 520 7: 43   
## 4:1869 4: 384 4:2702 4:1394 9: 3   
## 5:2924 5: 964 5:3514 5:4412   
## 6:2555 9: 125 6:4392 9:1161   
## 7:1281 9: 14   
## COPD EMPHY DEPRESS BRONCH SYMP.30D DUR.30D INCINDT   
## 1:2311 1: 973 1:4428 1:3198 66 :3094 12 :4181 1: 259   
## 2:9038 2:10425 2:7000 2:8159 30 :1989 6 :3094 2: 864   
## 7: 106 7: 61 7: 18 7: 100 100 :1425 10 :1425 3:10307   
## 9: 9 9: 5 9: 18 9: 7 88 : 613 1 :1149 7: 28   
## 2 : 468 2 : 804 9: 6   
## 3 : 454 11 : 613   
## (Other):3421 (Other): 198   
## LAST.MD LAST.MED LAST.SYMP EPIS.12M COMPASTH INS1   
## 4 :7077 1 :4479 1 :3161 1:4724 11 :3646 1:10833   
## 5 :1563 7 :2027 3 :2205 2:3565 6 :3094 2: 611   
## 6 : 673 3 :1388 7 :1669 6:3094 3 :2915 7: 14   
## 7 :2014 4 :1111 4 :1425 7: 77 1 :1060 9: 6   
## 77: 76 2 : 950 2 :1414 9: 4 2 : 703   
## 88: 55 5 : 931 5 : 924 7 : 33   
## 99: 6 (Other): 578 (Other): 666 (Other): 13   
## INS2 ER.VISIT HOSP.VST ASRXCOST WORKTALK  
## 1: 561 1:1229 1: 345 1:1393 1:2305   
## 2:10257 2:5896 2:5986 2:8279 2:8841   
## 5: 611 5:1772 4: 808 5:1772 6: 219   
## 7: 27 6:2546 5:1772 7: 13 7: 76   
## 9: 8 7: 20 6:2546 9: 7 8: 15   
## 9: 1 7: 7 9: 8   
##

#### Visualization of some combine variables



Proportion of Good Skill Management in terme of Education Level



Proportion of Good skill management in terme of Duration of Asthma Attack

#### Splitting the data into train and test sets

### BUILDS MODELS

#### Model using full predictors with glm

##   
## Call: glm(formula = TARGET ~ ., family = binomial, data = training1)  
##   
## Coefficients:  
## (Intercept) SEX2 AGEG.F72 AGEG.F73 AGEG.F74 AGEG.F75   
## -0.720247 -0.214534 -0.277247 -0.127407 -0.020848 0.287730   
## AGEG.F76 AGEG.F77 X\_RACEGR32 X\_RACEGR33 X\_RACEGR34 X\_RACEGR35   
## 0.465589 0.612869 0.006537 0.085871 -0.029290 -0.281095   
## X\_RACEGR39 EDUCAL2 EDUCAL3 EDUCAL4 EDUCAL5 EDUCAL6   
## 0.138785 -0.212176 -0.557089 -0.866679 -1.046080 -1.097045   
## EDUCAL9 X\_INCOMG2 X\_INCOMG3 X\_INCOMG4 X\_INCOMG5 X\_INCOMG9   
## -2.328621 -0.091065 -0.125717 -0.033362 -0.262513 0.047963   
## X\_RFBMI52 X\_RFBMI59 SMOKE1002 SMOKE1007 SMOKE1009 COPD2   
## 0.062864 0.058071 -0.056524 -0.757065 -13.750625 0.121718   
## COPD7 COPD9 EMPHY2 EMPHY7 EMPHY9 DEPRESS2   
## 0.629598 0.367690 -0.134204 -0.558472 0.340025 -0.014224   
## DEPRESS7 DEPRESS9 BRONCH2 BRONCH7 BRONCH9 SYMP.30D10   
## 0.009196 -0.047564 0.093950 0.260919 -1.447976 0.108451   
## SYMP.30D100 SYMP.30D11 SYMP.30D12 SYMP.30D13 SYMP.30D14 SYMP.30D15   
## -0.465675 -0.715201 -0.701890 -13.506496 0.022881 0.060340   
## SYMP.30D16 SYMP.30D17 SYMP.30D18 SYMP.30D19 SYMP.30D2 SYMP.30D20   
## 0.455604 0.184638 -0.134796 -13.996583 0.113954 0.163732   
## SYMP.30D21 SYMP.30D22 SYMP.30D23 SYMP.30D24 SYMP.30D25 SYMP.30D26   
## 0.015514 0.516111 0.257715 -0.184180 0.182948 1.125352   
## SYMP.30D27 SYMP.30D28 SYMP.30D29 SYMP.30D3 SYMP.30D30 SYMP.30D4   
## -0.358165 -0.323607 0.310424 0.019397 0.175445 0.092467   
## SYMP.30D5 SYMP.30D6 SYMP.30D66 SYMP.30D7 SYMP.30D77 SYMP.30D8   
## -0.031170 0.194517 0.881815 0.036005 0.160622 -0.248312   
## SYMP.30D88 SYMP.30D9 SYMP.30D99 DUR.30D10 DUR.30D11 DUR.30D12   
## 0.089904 0.146869 -1.596238 NA NA NA   
## DUR.30D2 DUR.30D6 DUR.30D7 DUR.30D77 DUR.30D9 DUR.30D99   
## 0.262947 NA 0.869861 NA 0.615843 NA   
## INCINDT2 INCINDT3 INCINDT7 INCINDT9 LAST.MD5 LAST.MD6   
## -0.661786 -1.031877 -0.271496 -0.156390 0.013633 0.293570   
## LAST.MD7 LAST.MD77 LAST.MD88 LAST.MD99 LAST.MED2 LAST.MED3   
## 0.493551 0.935354 0.895384 1.340885 0.234425 0.515061   
## LAST.MED4 LAST.MED5 LAST.MED6 LAST.MED7 LAST.MED77 LAST.MED99   
## 0.570780 0.495779 0.346164 0.286189 0.589096 40.010731   
## LAST.SYMP2 LAST.SYMP3 LAST.SYMP4 LAST.SYMP5 LAST.SYMP6 LAST.SYMP7   
## -0.100103 -0.267564 NA -0.830316 -0.921529 -1.032972   
## LAST.SYMP77 LAST.SYMP88 LAST.SYMP99 EPIS.12M2 EPIS.12M6 EPIS.12M7   
## 0.267289 NA -13.507718 0.547602 NA 1.310843   
## EPIS.12M9 COMPASTH11 COMPASTH2 COMPASTH3 COMPASTH4 COMPASTH6   
## -12.004831 NA 0.327608 0.326907 -0.633537 NA   
## COMPASTH7 COMPASTH9 INS12 INS17 INS19 INS22   
## 1.087720 -12.656774 0.239031 -1.298070 13.394431 0.058179   
## INS25 INS27 INS29 ER.VISIT2 ER.VISIT5 ER.VISIT6   
## NA 0.778523 -14.653787 0.009055 0.886264 0.909729   
## ER.VISIT7 ER.VISIT9 HOSP.VST2 HOSP.VST4 HOSP.VST5 HOSP.VST6   
## 0.343131 -13.567788 0.526691 0.543525 NA NA   
## HOSP.VST7 ASRXCOST2 ASRXCOST5 ASRXCOST7 ASRXCOST9 WORKTALK2   
## -16.262219 -0.092710 NA 0.691750 0.233892 0.793452   
## WORKTALK6 WORKTALK7 WORKTALK8 WORKTALK9   
## 0.766222 0.512975 1.109674 0.390355   
##   
## Degrees of Freedom: 9171 Total (i.e. Null); 9039 Residual  
## Null Deviance: 11490   
## Residual Deviance: 10510 AIC: 10780

#### Confusion Matrix with the testingset

#### First glm model using backward elimination of step function

##   
## Call: glm(formula = TARGET ~ SEX + AGEG.F7 + X\_RACEGR3 + EDUCAL + X\_INCOMG +   
## SMOKE100 + COPD + DUR.30D + INCINDT + LAST.MD + LAST.MED +   
## LAST.SYMP + EPIS.12M + COMPASTH + HOSP.VST + WORKTALK, family = binomial,   
## data = training1)  
##   
## Coefficients:  
## (Intercept) SEX2 AGEG.F72 AGEG.F73 AGEG.F74 AGEG.F75   
## -0.589332 -0.222979 -0.266369 -0.125935 -0.015489 0.286737   
## AGEG.F76 AGEG.F77 X\_RACEGR32 X\_RACEGR33 X\_RACEGR34 X\_RACEGR35   
## 0.451439 0.589598 0.015299 0.098344 -0.037626 -0.271327   
## X\_RACEGR39 EDUCAL2 EDUCAL3 EDUCAL4 EDUCAL5 EDUCAL6   
## 0.122686 -0.191693 -0.534684 -0.843293 -1.024266 -1.082214   
## EDUCAL9 X\_INCOMG2 X\_INCOMG3 X\_INCOMG4 X\_INCOMG5 X\_INCOMG9   
## -2.317873 -0.077383 -0.115368 -0.031883 -0.269049 0.050315   
## SMOKE1002 SMOKE1007 SMOKE1009 COPD2 COPD7 COPD9   
## -0.064508 -0.764376 -13.777105 0.116329 0.612199 0.053562   
## DUR.30D10 DUR.30D11 DUR.30D12 DUR.30D2 DUR.30D6 DUR.30D7   
## -0.661620 -0.084544 -0.130110 0.263353 0.691343 0.846535   
## DUR.30D77 DUR.30D9 DUR.30D99 INCINDT2 INCINDT3 INCINDT7   
## -0.009555 0.568977 -1.724367 -0.622656 -1.003880 -0.222372   
## INCINDT9 LAST.MD5 LAST.MD6 LAST.MD7 LAST.MD77 LAST.MD88   
## -0.123629 0.001754 0.292535 0.479780 0.925610 0.892171   
## LAST.MD99 LAST.MED2 LAST.MED3 LAST.MED4 LAST.MED5 LAST.MED6   
## 1.355357 0.235161 0.510217 0.568716 0.484293 0.350247   
## LAST.MED7 LAST.MED77 LAST.MED99 LAST.SYMP2 LAST.SYMP3 LAST.SYMP4   
## 0.281585 0.606270 38.816067 -0.109101 -0.286406 NA   
## LAST.SYMP5 LAST.SYMP6 LAST.SYMP7 LAST.SYMP77 LAST.SYMP88 LAST.SYMP99   
## -0.834919 -0.940326 -1.035606 0.255290 NA -13.550830   
## EPIS.12M2 EPIS.12M6 EPIS.12M7 EPIS.12M9 COMPASTH11 COMPASTH2   
## 0.544969 NA 1.318921 -12.090232 NA 0.344572   
## COMPASTH3 COMPASTH4 COMPASTH6 COMPASTH7 COMPASTH9 HOSP.VST2   
## 0.325250 -0.549496 NA 1.057530 -12.635818 0.531119   
## HOSP.VST4 HOSP.VST5 HOSP.VST6 HOSP.VST7 WORKTALK2 WORKTALK6   
## 0.553344 0.982279 0.924766 -14.575720 0.794086 0.757452   
## WORKTALK7 WORKTALK8 WORKTALK9   
## 0.506774 0.578476 0.332273   
##   
## Degrees of Freedom: 9171 Total (i.e. Null); 9090 Residual  
## Null Deviance: 11490   
## Residual Deviance: 10550 AIC: 10710

Call: glm(formula = TARGET ~ SEX + AGEG.F7 + X\_RACEGR3 + EDUCAL + BRONCH + DUR.30D + INCINDT + LAST.MD + LAST.MED + LAST.SYMP + COMPASTH + HOSPTIME + ASRXCOST + WORKTALK, family = binomial, data = training1)

#### Confusion Matrix with the testingset

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 1440 567  
## 1 119 166  
##   
## Accuracy : 0.7007   
## 95% CI : (0.6815, 0.7194)  
## No Information Rate : 0.6802   
## P-Value [Acc > NIR] : 0.01822   
##   
## Kappa : 0.1791   
##   
## Mcnemar's Test P-Value : < 2e-16   
##   
## Sensitivity : 0.22647   
## Specificity : 0.92367   
## Pos Pred Value : 0.58246   
## Neg Pred Value : 0.71749   
## Prevalence : 0.31981   
## Detection Rate : 0.07243   
## Detection Prevalence : 0.12435   
## Balanced Accuracy : 0.57507   
##   
## 'Positive' Class : 1   
##

#### Second glm model

##   
## Call: glm(formula = TARGET ~ SEX + AGEG.F7 + X\_RACEGR3 + EDUCAL + X\_INCOMG +   
## BRONCH + DUR.30D + INCINDT + LAST.MD + LAST.MED + LAST.SYMP +   
## COMPASTH + WORKTALK, family = binomial, data = training1)  
##   
## Coefficients:  
## (Intercept) SEX2 AGEG.F72 AGEG.F73 AGEG.F74 AGEG.F75   
## -0.192459 -0.223205 -0.253278 -0.110517 0.003903 0.303458   
## AGEG.F76 AGEG.F77 X\_RACEGR32 X\_RACEGR33 X\_RACEGR34 X\_RACEGR35   
## 0.470740 0.597405 0.001686 0.113525 -0.041497 -0.269093   
## X\_RACEGR39 EDUCAL2 EDUCAL3 EDUCAL4 EDUCAL5 EDUCAL6   
## 0.102469 -0.158920 -0.488902 -0.792705 -0.974917 -1.036767   
## EDUCAL9 X\_INCOMG2 X\_INCOMG3 X\_INCOMG4 X\_INCOMG5 X\_INCOMG9   
## -2.249575 -0.074492 -0.110922 -0.026521 -0.266633 0.052386   
## BRONCH2 BRONCH7 BRONCH9 DUR.30D10 DUR.30D11 DUR.30D12   
## 0.096854 0.309162 -0.870527 -0.656699 -0.062246 -0.112605   
## DUR.30D2 DUR.30D6 DUR.30D7 DUR.30D77 DUR.30D9 DUR.30D99   
## 0.277606 0.723253 0.864001 0.026642 0.612289 -1.639910   
## INCINDT2 INCINDT3 INCINDT7 INCINDT9 LAST.MD5 LAST.MD6   
## -0.601401 -0.993069 -0.407455 -0.108022 0.412143 0.698841   
## LAST.MD7 LAST.MD77 LAST.MD88 LAST.MD99 LAST.MED2 LAST.MED3   
## 0.887405 0.988821 1.335412 1.338723 0.250141 0.539728   
## LAST.MED4 LAST.MED5 LAST.MED6 LAST.MED7 LAST.MED77 LAST.MED99   
## 0.594971 0.512126 0.383665 0.318034 0.597444 24.517990   
## LAST.SYMP2 LAST.SYMP3 LAST.SYMP4 LAST.SYMP5 LAST.SYMP6 LAST.SYMP7   
## -0.107016 -0.295462 NA -0.825071 -0.932757 -1.017957   
## LAST.SYMP77 LAST.SYMP88 LAST.SYMP99 COMPASTH11 COMPASTH2 COMPASTH3   
## 0.230758 NA -12.746985 0.564527 0.318467 0.321465   
## COMPASTH4 COMPASTH6 COMPASTH7 COMPASTH9 WORKTALK2 WORKTALK6   
## -0.565743 NA 1.005412 -11.637517 0.795236 0.751750   
## WORKTALK7 WORKTALK8 WORKTALK9   
## 0.484428 0.290248 0.434927   
##   
## Degrees of Freedom: 9171 Total (i.e. Null); 9100 Residual  
## Null Deviance: 11490   
## Residual Deviance: 10580 AIC: 10730

#### Confusion Matrix with the testingset

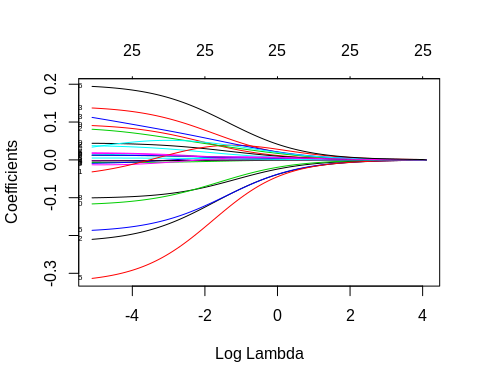
## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 1444 570  
## 1 115 163  
##   
## Accuracy : 0.7011   
## 95% CI : (0.6819, 0.7198)  
## No Information Rate : 0.6802   
## P-Value [Acc > NIR] : 0.01629   
##   
## Kappa : 0.1779   
##   
## Mcnemar's Test P-Value : < 2e-16   
##   
## Sensitivity : 0.22237   
## Specificity : 0.92623   
## Pos Pred Value : 0.58633   
## Neg Pred Value : 0.71698   
## Prevalence : 0.31981   
## Detection Rate : 0.07112   
## Detection Prevalence : 0.12129   
## Balanced Accuracy : 0.57430   
##   
## 'Positive' Class : 1   
##

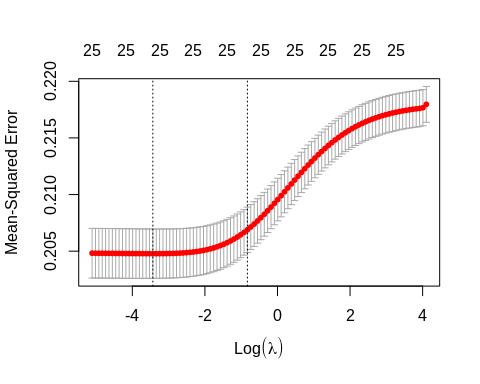
#### Lasso and Ridge model

Since our dataset has multiple variable, we can use penalized logistic regression to find an optimal performing model. Ridge Regression and Lasso Regression have two different approaches. Ridge Regression incorporates all variables in the model and gives the coefficients of variables with minor contribution close to zero Lasso Regression keeps only the most significant variables and gives zero to the coefficient of the rest of variables.

#### Split the data into trainset and testingset, Dumy code categorical predictors

#### Ridge Regression

We fit and obsrve the coefficients of rigde regression against the log of lambda.  The coefficients are significative for negative log lambda and start stabilize around -4



Lambda that Minimises MSE

The plot shows that the log of the optimal value of lambda (i.e. the one that minimises the root mean square error) is approximately -3. The exact value can be viewed by examining the variable lambda\_min in the code below. In general though, the objective of regularisation is to balance accuracy and simplicity. In the present context, this means a model with the smallest number of coefficients that also gives a good accuracy. To this end, the cv.glmnet function finds the value of lambda that gives the simplest model but also lies within one standard error of the optimal value of lambda.

## [1] 0.03224588

#### Confusion matrix with lambda min

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 1557 702  
## 1 14 19  
##   
## Accuracy : 0.6876   
## 95% CI : (0.6682, 0.7066)  
## No Information Rate : 0.6854   
## P-Value [Acc > NIR] : 0.4208   
##   
## Kappa : 0.0235   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.02635   
## Specificity : 0.99109   
## Pos Pred Value : 0.57576   
## Neg Pred Value : 0.68924   
## Prevalence : 0.31457   
## Detection Rate : 0.00829   
## Detection Prevalence : 0.01440   
## Balanced Accuracy : 0.50872   
##   
## 'Positive' Class : 1   
##

We observe overfitting with this ridge model

#### Confusion matrix with best lambda

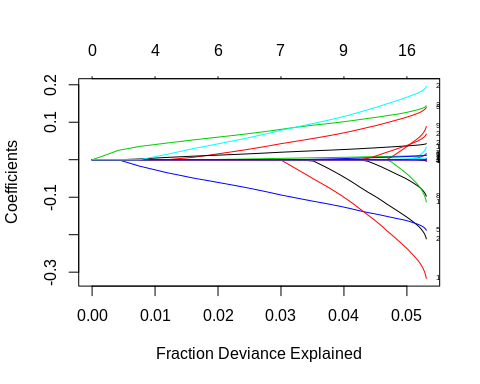
## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 1571 720  
## 1 0 1  
##   
## Accuracy : 0.6859   
## 95% CI : (0.6664, 0.7048)  
## No Information Rate : 0.6854   
## P-Value [Acc > NIR] : 0.4921   
##   
## Kappa : 0.0019   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.0013870   
## Specificity : 1.0000000   
## Pos Pred Value : 1.0000000   
## Neg Pred Value : 0.6857268   
## Prevalence : 0.3145724   
## Detection Rate : 0.0004363   
## Detection Prevalence : 0.0004363   
## Balanced Accuracy : 0.5006935   
##   
## 'Positive' Class : 1   
##

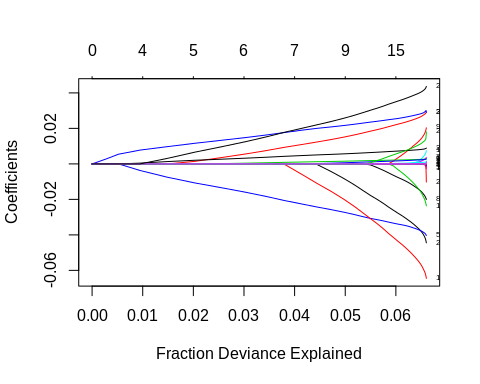
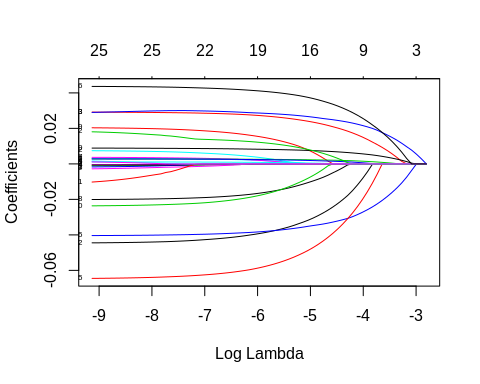
We observe overfitting with this second ridge model

#### Getting the coefficients

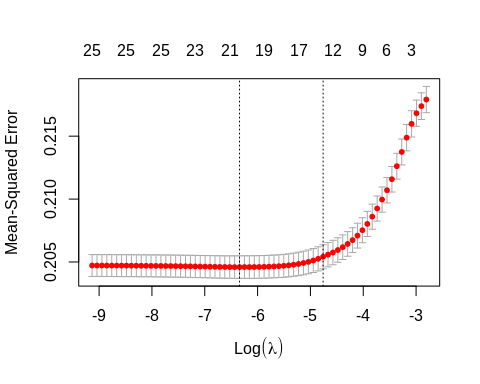
## 27 x 1 sparse Matrix of class "dgCMatrix"  
## s0  
## (Intercept) -0.547811093  
## (Intercept) .   
## SEX -0.183135695  
## AGEG.F7 0.118738150  
## X\_RACEGR3 -0.006391591  
## EDUCAL -0.167341239  
## X\_INCOMG 0.011981641  
## X\_RFBMI5 0.015681273  
## SMOKE100 -0.093582336  
## COPD 0.075708623  
## EMPHY -0.102295270  
## DEPRESS -0.003255346  
## BRONCH 0.031700069  
## SYMP.30D -0.001502735  
## DUR.30D -0.002133493  
## INCINDT -0.270916826  
## LAST.MD 0.011127531  
## LAST.MED 0.011135219  
## LAST.SYMP 0.004511173  
## EPIS.12M -0.009088852  
## COMPASTH 0.040055398  
## INS1 0.002107476  
## INS2 0.064283834  
## ER.VISIT 0.084718679  
## HOSP.VST 0.050487532  
## ASRXCOST 0.013981691  
## WORKTALK 0.175028675

##### Lasso Regression





#### Find the best lambda using cross validation

 The plot shows that the log of the optimal value of lambda (i.e. the one that minimises the root mean square error) is approximately -10. The exact value can be viewed by examining the variable lambda\_min in the code below. In general though, the objective of regularisation is to balance accuracy and simplicity. In the present context, this means a model with the smallest number of coefficients that also gives a good accuracy. To this end, the cv.glmnet function finds the value of lambda that gives the simplest model but also lies within one standard error of the optimal value of lambda.

#### Confusion Matrix with lambda min

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 1490 637  
## 1 81 84  
##   
## Accuracy : 0.6867   
## 95% CI : (0.6673, 0.7057)  
## No Information Rate : 0.6854   
## P-Value [Acc > NIR] : 0.4563   
##   
## Kappa : 0.0821   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.11650   
## Specificity : 0.94844   
## Pos Pred Value : 0.50909   
## Neg Pred Value : 0.70052   
## Prevalence : 0.31457   
## Detection Rate : 0.03665   
## Detection Prevalence : 0.07199   
## Balanced Accuracy : 0.53247   
##   
## 'Positive' Class : 1   
##

#### Getting the coefficients

## 27 x 1 sparse Matrix of class "dgCMatrix"  
## s0  
## (Intercept) -0.486095629  
## (Intercept) .   
## SEX -0.199596041  
## AGEG.F7 0.134531746  
## X\_RACEGR3 -0.001252628  
## EDUCAL -0.182517500  
## X\_INCOMG 0.011776022  
## X\_RFBMI5 0.013102866  
## SMOKE100 -0.088933624  
## COPD 0.078096807  
## EMPHY -0.097855786  
## DEPRESS .   
## BRONCH 0.025530114  
## SYMP.30D -0.001503286  
## DUR.30D -0.001857234  
## INCINDT -0.301042143  
## LAST.MD 0.011538669  
## LAST.MED 0.011359148  
## LAST.SYMP 0.004289324  
## EPIS.12M .   
## COMPASTH 0.041544908  
## INS1 .   
## INS2 0.062292138  
## ER.VISIT 0.140318957  
## HOSP.VST .   
## ASRXCOST .   
## WORKTALK 0.189677901

#### Confusion Matrix with best lambda

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 1528 669  
## 1 43 52  
##   
## Accuracy : 0.6894   
## 95% CI : (0.67, 0.7083)  
## No Information Rate : 0.6854   
## P-Value [Acc > NIR] : 0.352   
##   
## Kappa : 0.0585   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.07212   
## Specificity : 0.97263   
## Pos Pred Value : 0.54737   
## Neg Pred Value : 0.69549   
## Prevalence : 0.31457   
## Detection Rate : 0.02269   
## Detection Prevalence : 0.04145   
## Balanced Accuracy : 0.52238   
##   
## 'Positive' Class : 1   
##

##### Calculating the AICc of Ridge and Lasso Models

it <- glmnet(x, y, family = “multinomial”)

tLL <- fitdf n <- fit$nobs AICc <- -tLL+2*k+2*k\*(k+1)/(n-k-1) AICc

## [1] -550.842

## [1] -568.2678

#### Partial Least Squared

#### Confusion Matrix with best lambda

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 1456 586  
## 1 103 147  
##   
## Accuracy : 0.6994   
## 95% CI : (0.6802, 0.7181)  
## No Information Rate : 0.6802   
## P-Value [Acc > NIR] : 0.02523   
##   
## Kappa : 0.1629   
##   
## Mcnemar's Test P-Value : < 2e-16   
##   
## Sensitivity : 0.20055   
## Specificity : 0.93393   
## Pos Pred Value : 0.58800   
## Neg Pred Value : 0.71303   
## Prevalence : 0.31981   
## Detection Rate : 0.06414   
## Detection Prevalence : 0.10908   
## Balanced Accuracy : 0.56724   
##   
## 'Positive' Class : 1   
##

#### Here we train the model with partial least square using tune parameter.

## Partial Least Squares   
##   
## 9172 samples  
## 25 predictor  
## 2 classes: 'F', 'T'   
##   
## Pre-processing: centered (147), scaled (147)   
## Resampling: Cross-Validated (10 fold, repeated 3 times)   
## Summary of sample sizes: 8255, 8255, 8255, 8255, 8254, 8254, ...   
## Resampling results across tuning parameters:  
##   
## ncomp ROC Sens Spec   
## 1 0.6368401 0.9610577 0.1109625  
## 2 0.6512083 0.9450855 0.1515440  
## 3 0.6606066 0.9394765 0.1713214  
## 4 0.6625909 0.9365919 0.1795067  
## 5 0.6646051 0.9361111 0.1766645  
## 6 0.6670429 0.9363782 0.1846223  
## 7 0.6681666 0.9350427 0.1870164  
## 8 0.6682807 0.9342415 0.1878104  
## 9 0.6681805 0.9349359 0.1858795  
## 10 0.6683792 0.9354167 0.1858776  
## 11 0.6682949 0.9354701 0.1875817  
## 12 0.6686542 0.9347222 0.1884919  
## 13 0.6686172 0.9341880 0.1881502  
## 14 0.6685352 0.9349359 0.1879242  
## 15 0.6683908 0.9342949 0.1898582  
##   
## ROC was used to select the optimal model using the largest value.  
## The final value used for the model was ncomp = 12.

#### Confusion Matrix with best lambda

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 1456 586  
## 1 103 147  
##   
## Accuracy : 0.6994   
## 95% CI : (0.6802, 0.7181)  
## No Information Rate : 0.6802   
## P-Value [Acc > NIR] : 0.02523   
##   
## Kappa : 0.1629   
##   
## Mcnemar's Test P-Value : < 2e-16   
##   
## Sensitivity : 0.20055   
## Specificity : 0.93393   
## Pos Pred Value : 0.58800   
## Neg Pred Value : 0.71303   
## Prevalence : 0.31981   
## Detection Rate : 0.06414   
## Detection Prevalence : 0.10908   
## Balanced Accuracy : 0.56724   
##   
## 'Positive' Class : 1   
##

### SELECT MODELS

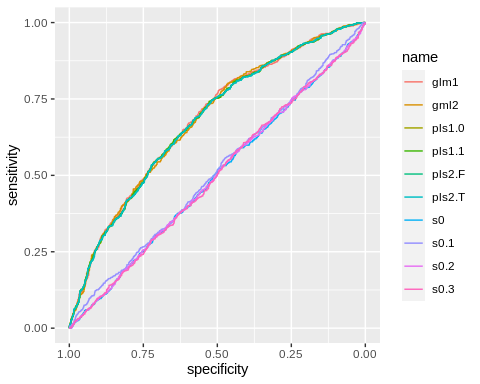
#### We compare the models with the accuray, precision, sensitivity, specificity, and F1 score from the confusion matrix

## glm.mod11 glm.mod12 ridge.mod1 ridge.mod2 lasso.mod1 lasso.mod2  
## Accuracy 0.7006981 0.7011344 0.68760908 0.685863874 0.6867365 0.68935428  
## Precision 0.5824561 0.5863309 0.57575758 1.000000000 0.5090909 0.54736842  
## Sensitivity 0.2264666 0.2223738 0.02635229 0.001386963 0.1165049 0.07212205  
## Specificity 0.9236690 0.9262348 0.99108848 1.000000000 0.9484405 0.97262890  
## F1 0.3261297 0.3224530 0.05039788 0.002770083 0.1896163 0.12745098  
## pls.mod1 pls.mod2  
## Accuracy 0.6993892 0.6993892  
## Precision 0.5880000 0.5880000  
## Sensitivity 0.2005457 0.2005457  
## Specificity 0.9339320 0.9339320  
## F1 0.2990844 0.2990844

With precision and specificity equal to 1, the ridge.mod2 model is overfitting. But glm.mod12 has the best accuracy, precision, sensivity, and specificity.

### Using pROC package.

We can plot the ROC curve and extract the AUC value.

 The glm model has the best Area Under the Curve.

### We run the glm model with the entire dataset

#### Second glm model

##   
## Call:  
## glm(formula = TARGET ~ SEX + AGEG.F7 + X\_RACEGR3 + EDUCAL + X\_INCOMG +   
## BRONCH + DUR.30D + INCINDT + LAST.MD + LAST.MED + LAST.SYMP +   
## COMPASTH + WORKTALK, family = binomial, data = asth.mgt.ad.min35)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.9490 -0.8763 -0.6594 1.1764 2.4806   
##   
## Coefficients: (3 not defined because of singularities)  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.38210 0.60604 -0.630 0.528370   
## SEX2 -0.23242 0.04649 -4.999 5.76e-07 \*\*\*  
## AGEG.F72 -0.16370 0.11715 -1.397 0.162308   
## AGEG.F73 -0.02776 0.11398 -0.244 0.807575   
## AGEG.F74 0.10525 0.10737 0.980 0.326942   
## AGEG.F75 0.37324 0.10223 3.651 0.000261 \*\*\*  
## AGEG.F76 0.55106 0.10379 5.309 1.10e-07 \*\*\*  
## AGEG.F77 0.76483 0.11162 6.852 7.27e-12 \*\*\*  
## X\_RACEGR32 -0.03110 0.09493 -0.328 0.743207   
## X\_RACEGR33 0.17165 0.11091 1.548 0.121697   
## X\_RACEGR34 -0.08139 0.12062 -0.675 0.499849   
## X\_RACEGR35 -0.19226 0.08278 -2.323 0.020199 \*   
## X\_RACEGR39 0.06120 0.20172 0.303 0.761610   
## EDUCAL2 -0.16482 0.58368 -0.282 0.777656   
## EDUCAL3 -0.43169 0.57744 -0.748 0.454706   
## EDUCAL4 -0.73311 0.57311 -1.279 0.200831   
## EDUCAL5 -0.89173 0.57336 -1.555 0.119878   
## EDUCAL6 -0.93515 0.57383 -1.630 0.103175   
## EDUCAL9 -2.26597 0.98212 -2.307 0.021043 \*   
## X\_INCOMG2 -0.03556 0.07704 -0.461 0.644442   
## X\_INCOMG3 -0.09969 0.09132 -1.092 0.275011   
## X\_INCOMG4 -0.04667 0.08512 -0.548 0.583532   
## X\_INCOMG5 -0.25447 0.07386 -3.445 0.000570 \*\*\*  
## X\_INCOMG9 0.12638 0.08631 1.464 0.143135   
## BRONCH2 0.09994 0.05106 1.957 0.050290 .   
## BRONCH7 0.10309 0.22628 0.456 0.648692   
## BRONCH9 -1.02128 1.11338 -0.917 0.359000   
## DUR.30D10 -0.61824 0.10348 -5.975 2.31e-09 \*\*\*  
## DUR.30D11 -0.15718 0.13206 -1.190 0.233951   
## DUR.30D12 -0.10548 0.08549 -1.234 0.217288   
## DUR.30D2 0.20167 0.10189 1.979 0.047782 \*   
## DUR.30D6 0.64787 0.32579 1.989 0.046746 \*   
## DUR.30D7 0.72773 0.37434 1.944 0.051892 .   
## DUR.30D77 0.01551 0.19536 0.079 0.936727   
## DUR.30D9 0.57110 1.41810 0.403 0.687154   
## DUR.30D99 -1.93842 0.93956 -2.063 0.039102 \*   
## INCINDT2 -0.56605 0.15138 -3.739 0.000185 \*\*\*  
## INCINDT3 -0.99069 0.13484 -7.347 2.03e-13 \*\*\*  
## INCINDT7 -0.48666 0.44205 -1.101 0.270937   
## INCINDT9 -0.40101 0.87094 -0.460 0.645200   
## LAST.MD5 0.39121 0.06592 5.935 2.95e-09 \*\*\*  
## LAST.MD6 0.62743 0.09556 6.566 5.17e-11 \*\*\*  
## LAST.MD7 0.86709 0.08337 10.400 < 2e-16 \*\*\*  
## LAST.MD77 1.14999 0.25118 4.578 4.69e-06 \*\*\*  
## LAST.MD88 1.29113 0.29610 4.360 1.30e-05 \*\*\*  
## LAST.MD99 0.67928 1.00919 0.673 0.500884   
## LAST.MED2 0.32132 0.08600 3.736 0.000187 \*\*\*  
## LAST.MED3 0.53869 0.07766 6.936 4.03e-12 \*\*\*  
## LAST.MED4 0.56081 0.08918 6.289 3.20e-10 \*\*\*  
## LAST.MED5 0.52468 0.09717 5.400 6.68e-08 \*\*\*  
## LAST.MED6 0.45120 0.12629 3.573 0.000353 \*\*\*  
## LAST.MED7 0.43085 0.10447 4.124 3.72e-05 \*\*\*  
## LAST.MED77 0.48773 0.26304 1.854 0.063711 .   
## LAST.MED99 25.08368 241.45581 0.104 0.917260   
## LAST.SYMP2 -0.11672 0.08210 -1.422 0.155107   
## LAST.SYMP3 -0.28507 0.08042 -3.545 0.000393 \*\*\*  
## LAST.SYMP4 NA NA NA NA   
## LAST.SYMP5 -0.78003 0.31769 -2.455 0.014075 \*   
## LAST.SYMP6 -0.88437 0.32940 -2.685 0.007258 \*\*   
## LAST.SYMP7 -1.04953 0.31794 -3.301 0.000963 \*\*\*  
## LAST.SYMP77 -0.01259 0.19094 -0.066 0.947426   
## LAST.SYMP88 NA NA NA NA   
## LAST.SYMP99 -12.77092 149.86193 -0.085 0.932088   
## COMPASTH11 0.53268 0.08687 6.132 8.70e-10 \*\*\*  
## COMPASTH2 0.34129 0.11807 2.891 0.003844 \*\*   
## COMPASTH3 0.28092 0.08942 3.141 0.001681 \*\*   
## COMPASTH4 0.22616 0.62366 0.363 0.716879   
## COMPASTH6 NA NA NA NA   
## COMPASTH7 0.97156 0.37276 2.606 0.009150 \*\*   
## COMPASTH9 -11.72182 535.41120 -0.022 0.982533   
## WORKTALK2 0.82307 0.06096 13.501 < 2e-16 \*\*\*  
## WORKTALK6 0.84767 0.15960 5.311 1.09e-07 \*\*\*  
## WORKTALK7 0.71938 0.25734 2.795 0.005183 \*\*   
## WORKTALK8 0.15003 0.62846 0.239 0.811318   
## WORKTALK9 1.38387 0.80164 1.726 0.084294 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 14368 on 11463 degrees of freedom  
## Residual deviance: 13233 on 11392 degrees of freedom  
## AIC: 13377  
##   
## Number of Fisher Scoring iterations: 12

## (Intercept) SEX2 AGEG.F72 AGEG.F73 AGEG.F74 AGEG.F75   
## 6.824238e-01 7.926169e-01 8.489942e-01 9.726218e-01 1.110988e+00 1.452435e+00   
## AGEG.F76 AGEG.F77 X\_RACEGR32 X\_RACEGR33 X\_RACEGR34 X\_RACEGR35   
## 1.735084e+00 2.148628e+00 9.693780e-01 1.187263e+00 9.218374e-01 8.250903e-01   
## X\_RACEGR39 EDUCAL2 EDUCAL3 EDUCAL4 EDUCAL5 EDUCAL6   
## 1.063107e+00 8.480487e-01 6.494101e-01 4.804127e-01 4.099451e-01 3.925281e-01   
## EDUCAL9 X\_INCOMG2 X\_INCOMG3 X\_INCOMG4 X\_INCOMG5 X\_INCOMG9   
## 1.037289e-01 9.650689e-01 9.051216e-01 9.544046e-01 7.753252e-01 1.134711e+00   
## BRONCH2 BRONCH7 BRONCH9 DUR.30D10 DUR.30D11 DUR.30D12   
## 1.105109e+00 1.108592e+00 3.601350e-01 5.388918e-01 8.545486e-01 8.998968e-01   
## DUR.30D2 DUR.30D6 DUR.30D7 DUR.30D77 DUR.30D9 DUR.30D99   
## 1.223443e+00 1.911470e+00 2.070367e+00 1.015629e+00 1.770211e+00 1.439309e-01   
## INCINDT2 INCINDT3 INCINDT7 INCINDT9 LAST.MD5 LAST.MD6   
## 5.677622e-01 3.713205e-01 6.146765e-01 6.696403e-01 1.478775e+00 1.872793e+00   
## LAST.MD7 LAST.MD77 LAST.MD88 LAST.MD99 LAST.MED2 LAST.MED3   
## 2.379968e+00 3.158162e+00 3.636883e+00 1.972465e+00 1.378949e+00 1.713766e+00   
## LAST.MED4 LAST.MED5 LAST.MED6 LAST.MED7 LAST.MED77 LAST.MED99   
## 1.752085e+00 1.689925e+00 1.570191e+00 1.538557e+00 1.628612e+00 7.828968e+10   
## LAST.SYMP2 LAST.SYMP3 LAST.SYMP4 LAST.SYMP5 LAST.SYMP6 LAST.SYMP7   
## 8.898343e-01 7.519638e-01 NA 4.583930e-01 4.129734e-01 3.501014e-01   
## LAST.SYMP77 LAST.SYMP88 LAST.SYMP99 COMPASTH11 COMPASTH2 COMPASTH3   
## 9.874886e-01 NA 2.842223e-06 1.703488e+00 1.406768e+00 1.324347e+00   
## COMPASTH4 COMPASTH6 COMPASTH7 COMPASTH9 WORKTALK2 WORKTALK6   
## 1.253775e+00 NA 2.642054e+00 8.114815e-06 2.277485e+00 2.334212e+00   
## WORKTALK7 WORKTALK8 WORKTALK9   
## 2.053151e+00 1.161869e+00 3.990311e+00