Stat 615 Final_Project

Dhruv Jain & mekdim

2023-04-20

ACKNOWLEDGEMENTS

"I am not what happened to me, I am what I choose to become" by Christopher Gardner, The Pursuit of Happiness.

It is always a pleasure to remember the fine people who guided me in the Regression program. I received to uphold my practical and theoretical skills during the respective session. Firstly, I would like to thank **Pro. James C. Dickens** and secondly, I want to thank my family & friends for their love, motivation, and support during this semester in American university. Thanks for all the ideas, opinions, knowledge, and suggestions given to me to help me to complete this report. We are very thankful to American University for giving us the opportunity to pursue this project.

Title Page with Executive Summary

Title: Estimating Medical Cost.

Type of analysis: Application analysis

Table 1:

Name	course
Dhruv Jain	STAT -615
Mekdim Ashebo	STAT -615

```r
# calling all the libraries used in the code book
library(olsrr)
library(tidyverse)
library(dbplyr)
library(dplyr)
library(Matrix)
library(MASS)
library(ggplot2)
library(tibble)
library(data.table)
library(ggmosaic)
library(ggforce)
library(ggmap)
library(ggthemes)
library(purrr)
library(keep)
library(readr)
library(gridExtra)
library(randomForest)
library(corrplot)
library(PerformanceAnalytics)
```

1

1.1 About Data set

```r
# offer a preliminary description of the data set. For example, indicate the size of the data source, d

#data set source: https://www.kaggle.com/datasets/mirichoi0218/insurance

# Columns Description

#age: age of primary beneficiary
#sex: insurance contractor gender, female, male
#bmi: Body mass index, providing an understanding of body, weights that are relatively high or low rela
#objective index of body weight (kg / m ^ 2) using the ratio of height to weight, ideally 18.5 to 24.9
#children: Number of children covered by health insurance / Number of dependents
#smoker: Smoking
#region: the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.
#charges: Individual medical costs billed by health insurance


# We had to randomly sample 300 rows from our original data.
# We then saved these 300 rows into csv file so that we can import them later.
# going forward We will take that csv file. (which has only be run once)

# The preliminary steps we did
#insurance <- read_csv('Downloads/insurance.csv')
#insurance_300 <- sample_n(insurance, 300)

#write.csv(insurance_300 , file = "Desktop/insurance_300.csv")
```

```r
# Let import the 300 rows

# insurance_new <- read_csv("insurance_300.csv")

# Read in CSV file and specify column types
insurance_new <- read_csv("insurance_300.csv",
col_types = cols(
  age = col_double(),
  sex = col_character(),
  bmi = col_double(),
  children = col_double(),
  smoker = col_character(),
  region = col_character(),
  charges = col_double()
))
#300 rows  and 7 columns
# This project is about determining the factors that affect medical costs billed by health insurance
# The independent variables include three categorical variables and three quantitative variables.
# Sex, region(Northeast, northwest etc), and smoker(whether a person smokes or not) are the categorical
# variables. While the quantitative variables include the BMI index, the age and the number of children

nrow(insurance_new)
```

```
## [1] 300
```

```r
ncol(insurance_new)
```

```
## [1] 7
```

```r
# Let us quickly investigate the summary of our dependent variable
#  The median insurance charge is around 10097 and the mean of 13283. The
# standard deviation is 11399.

summary(insurance_new$charges)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1136    5134   10097   13283   17154   51195
```

```r
sd(insurance_new$charges)
```

```
## [1] 11399.06
```

```r
head(insurance_new,10)
```

```
## # A tibble: 10 x 7
##      age sex      bmi children smoker region    charges
##    <dbl> <chr>  <dbl>    <dbl> <chr>  <chr>       <dbl>
##  1    63 female  25.1        0 no     northwest  14255.
##  2    18 male    38.2        0 yes    southeast  36308.
##  3    48 male    29.6        0 no     southwest  21232.
```

```
##  4    46 female  33.4        1 no     southeast   8241.
##  5    52 male    30.2        1 no     southwest   9725.
##  6    36 female  19.9        0 no     northeast   5458.
##  7    19 male    20.9        1 no     southwest   1832.
##  8    48 male    36.7        1 no     northwest  28469.
##  9    19 female  29.8        0 no     southwest   1744.
## 10    19 female  20.6        0 no     southwest   1732.
```

1.2 cleaning the data and type of columns

```
# calling the data set using read csv file
# insurance_new <- read_csv('insurance_300.csv')
# number of rows in data
nrow(insurance_new)
```

```
## [1] 300
```

```
# number of colums in data set
ncol(insurance_new)
```

```
## [1] 7
```

```
# colums names
colnames(insurance_new)
```

```
## [1] "age"      "sex"      "bmi"      "children" "smoker"   "region"   "charges"
```

```
# visual data set look like
head(insurance_new,10)
```

```
## # A tibble: 10 x 7
##      age sex      bmi children smoker region    charges
##    <dbl> <chr>  <dbl>    <dbl> <chr>  <chr>       <dbl>
##  1    63 female  25.1        0 no     northwest  14255.
##  2    18 male    38.2        0 yes    southeast  36308.
##  3    48 male    29.6        0 no     southwest  21232.
##  4    46 female  33.4        1 no     southeast   8241.
##  5    52 male    30.2        1 no     southwest   9725.
##  6    36 female  19.9        0 no     northeast   5458.
##  7    19 male    20.9        1 no     southwest   1832.
##  8    48 male    36.7        1 no     northwest  28469.
##  9    19 female  29.8        0 no     southwest   1744.
## 10    19 female  20.6        0 no     southwest   1732.
```

```
# type of columns used in data frame (double, charterer)
str(insurance_new)
```

```
## spc_tbl_ [300 x 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ age     : num [1:300] 63 18 48 46 52 36 19 48 19 19 ...
```

```
##  $ sex     : chr [1:300] "female" "male" "male" "female" ...
##  $ bmi     : num [1:300] 25.1 38.2 29.6 33.4 30.2 ...
##  $ children: num [1:300] 0 0 0 1 1 0 1 1 0 0 ...
##  $ smoker  : chr [1:300] "no" "yes" "no" "no" ...
##  $ region  : chr [1:300] "northwest" "southeast" "southwest" "southeast" ...
##  $ charges : num [1:300] 14255 36308 21232 8241 9725 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   age = col_double(),
##   ..   sex = col_character(),
##   ..   bmi = col_double(),
##   ..   children = col_double(),
##   ..   smoker = col_character(),
##   ..   region = col_character(),
##   ..   charges = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
# summary of data colum wise
summary(insurance_new)
```

```
##       age            sex                 bmi           children
##  Min.   :18.00   Length:300         Min.   :17.29   Min.   :0.00
##  1st Qu.:27.00   Class :character   1st Qu.:25.25   1st Qu.:0.00
##  Median :40.50   Mode  :character   Median :30.01   Median :1.00
##  Mean   :39.88                      Mean   :30.02   Mean   :1.02
##  3rd Qu.:53.00                      3rd Qu.:34.20   3rd Qu.:2.00
##  Max.   :64.00                      Max.   :46.75   Max.   :5.00
##     smoker             region            charges
##  Length:300         Length:300         Min.   : 1136
##  Class :character   Class :character   1st Qu.: 5134
##  Mode  :character   Mode  :character   Median :10097
##                                        Mean   :13283
##                                        3rd Qu.:17154
##                                        Max.   :51195
```

```
# calculating NA/missing data in columns
colSums(is.na(insurance_new))
```

```
##      age      sex      bmi children   smoker   region  charges
##        0        0        0        0        0        0        0
```

```
# converting to factor variable
insurance_new$sex = as.factor(insurance_new$sex)
insurance_new$smoker = as.factor(insurance_new$smoker)
# how many unique values
unique(insurance_new$sex)
```

```
## [1] female male
## Levels: female male
```

```r
unique(insurance_new$children)
```

```
## [1] 0 1 2 5 3 4
```

```r
unique(insurance_new$smoker)
```

```
## [1] no  yes
## Levels: no yes
```

```r
unique(insurance_new$region)
```

```
## [1] "northwest" "southeast" "southwest" "northeast"
```

```r
# Check levels of smoker variable
table(insurance_new$smoker)
```

```
##
##  no yes
## 239  61
```

```r
# Check levels of region variable
table(insurance_new$region)
```

```
##
## northeast northwest southeast southwest
##        67        77        76        80
```

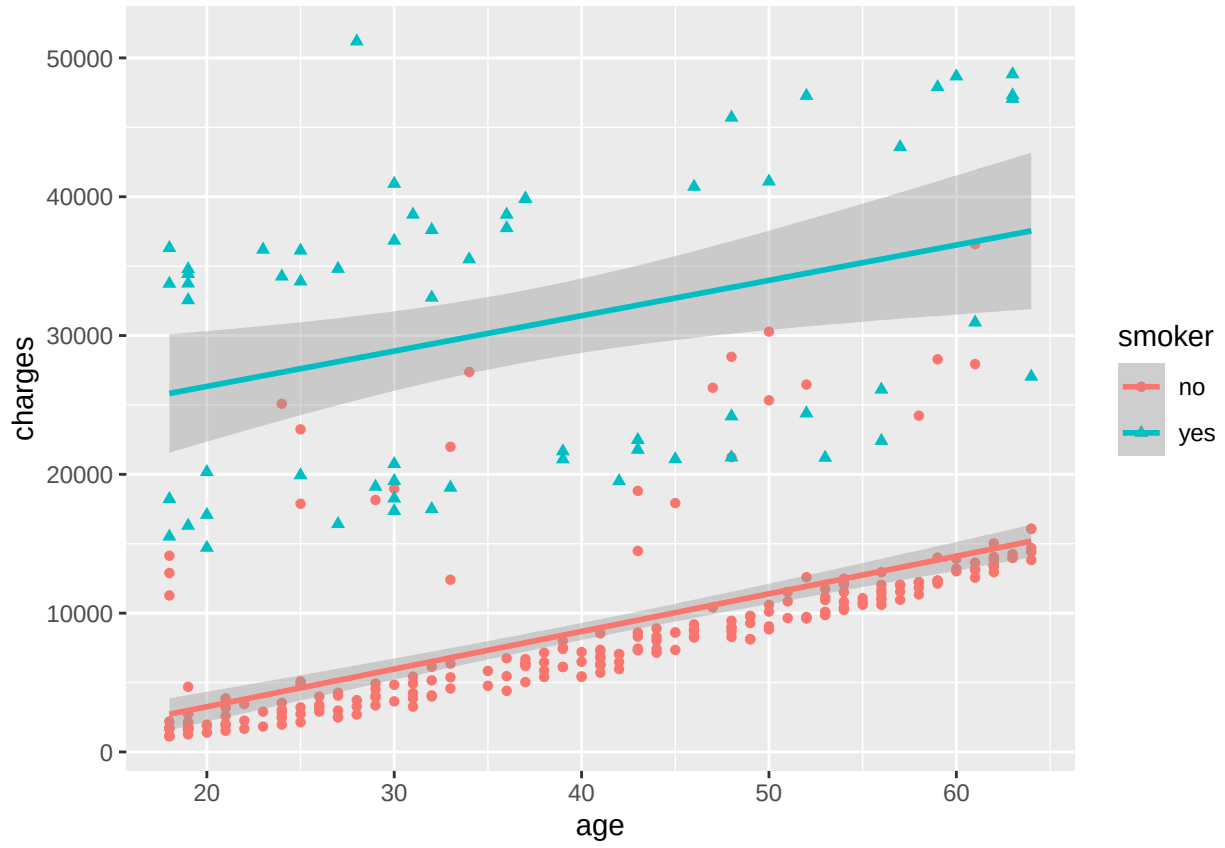
```r
# Check levels of sex variable
table(insurance_new$sex)
```

```
##
## female   male
##    134    166
```

1.3 visualization

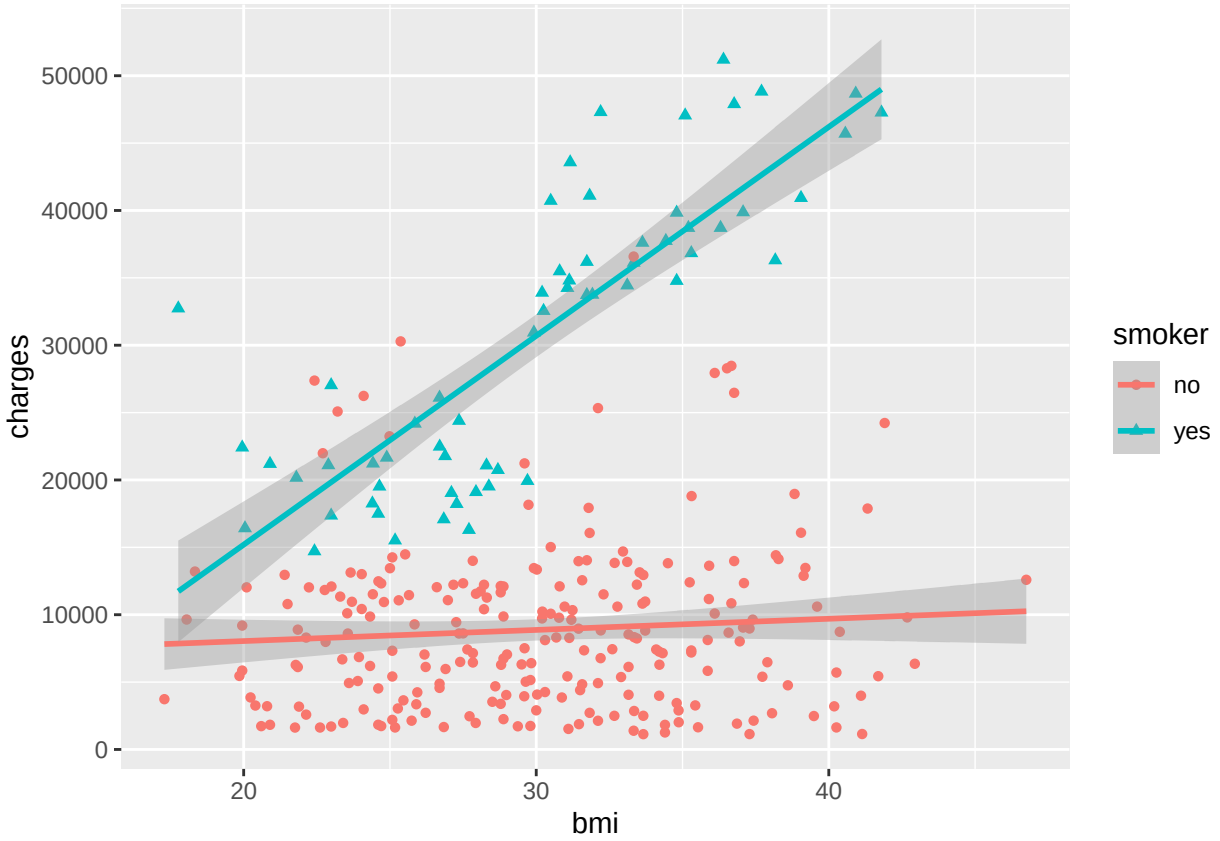
```r
# Does age affect medical charges for smoker?
ggplot(data = insurance_new ,
       aes(x=age, y=charges,shape=smoker,color = smoker)) +
  geom_point()+
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# Yes the charges are increased as we increase the number of age. Now the fun part is if a person smoke

# Does Body mass index (BMI) affect medical charges for smoker?
ggplot(data = insurance_new ,
       aes(x=bmi, y=charges,shape=smoker,color = smoker)) +
  geom_point()+
  geom_smooth(method=lm)
```

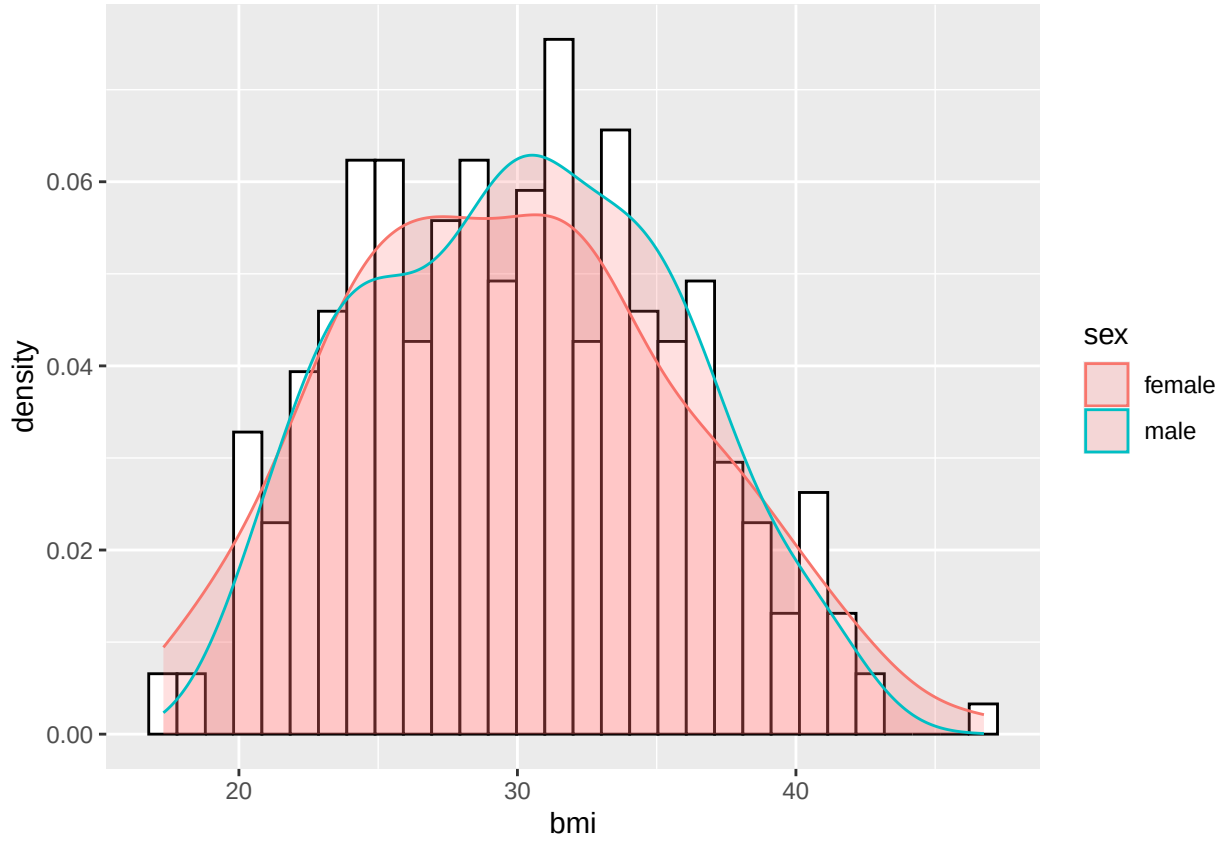
```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# One can clearly observe that smoking affect in BMI and increased  with the medical expenses.


# Histogram for density graph for Body mass index (BMI)
ggplot(data = insurance_new , aes(x=bmi,color=sex)) +
 geom_histogram(aes(y=..density..), colour="black", fill="white")+
 geom_density(alpha=.2, fill="#FF6666")
```

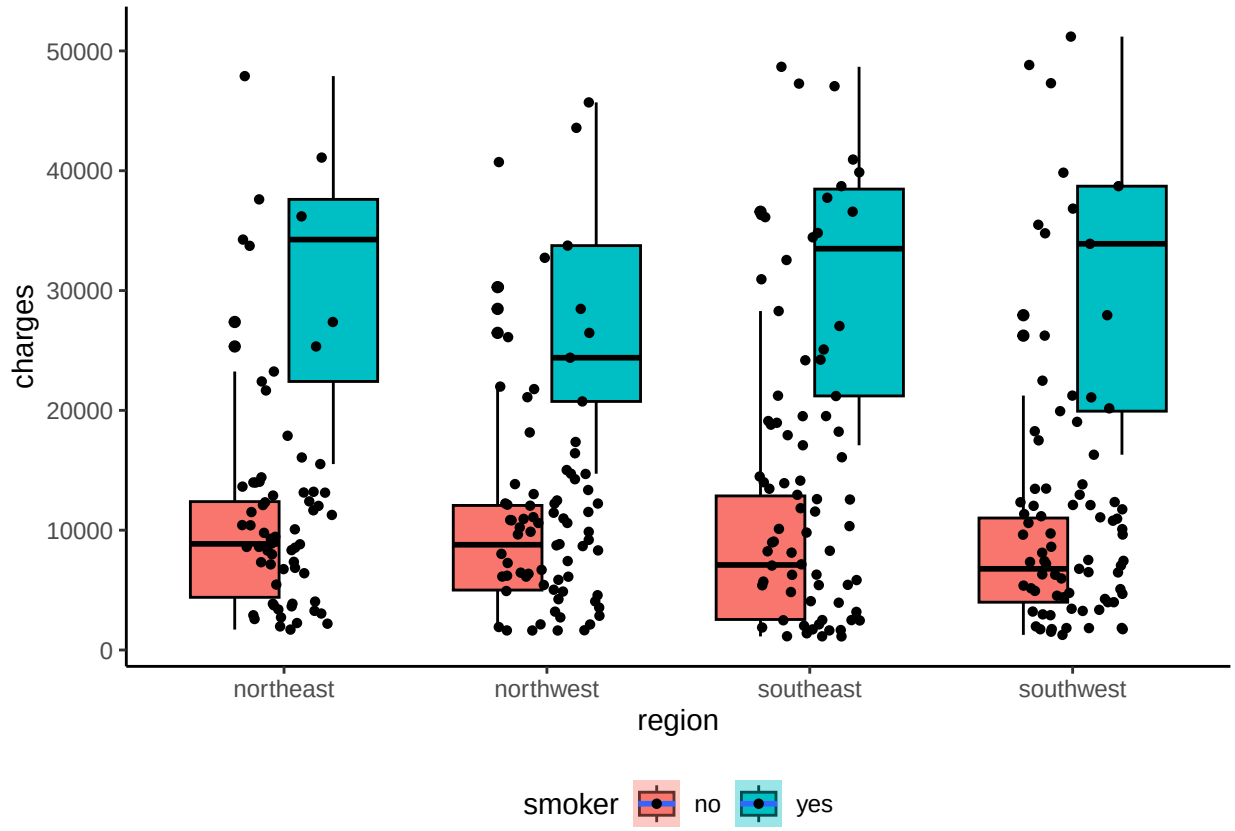
```
## Warning: The dot-dot notation (‘..density..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(density)‘ instead.
```

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



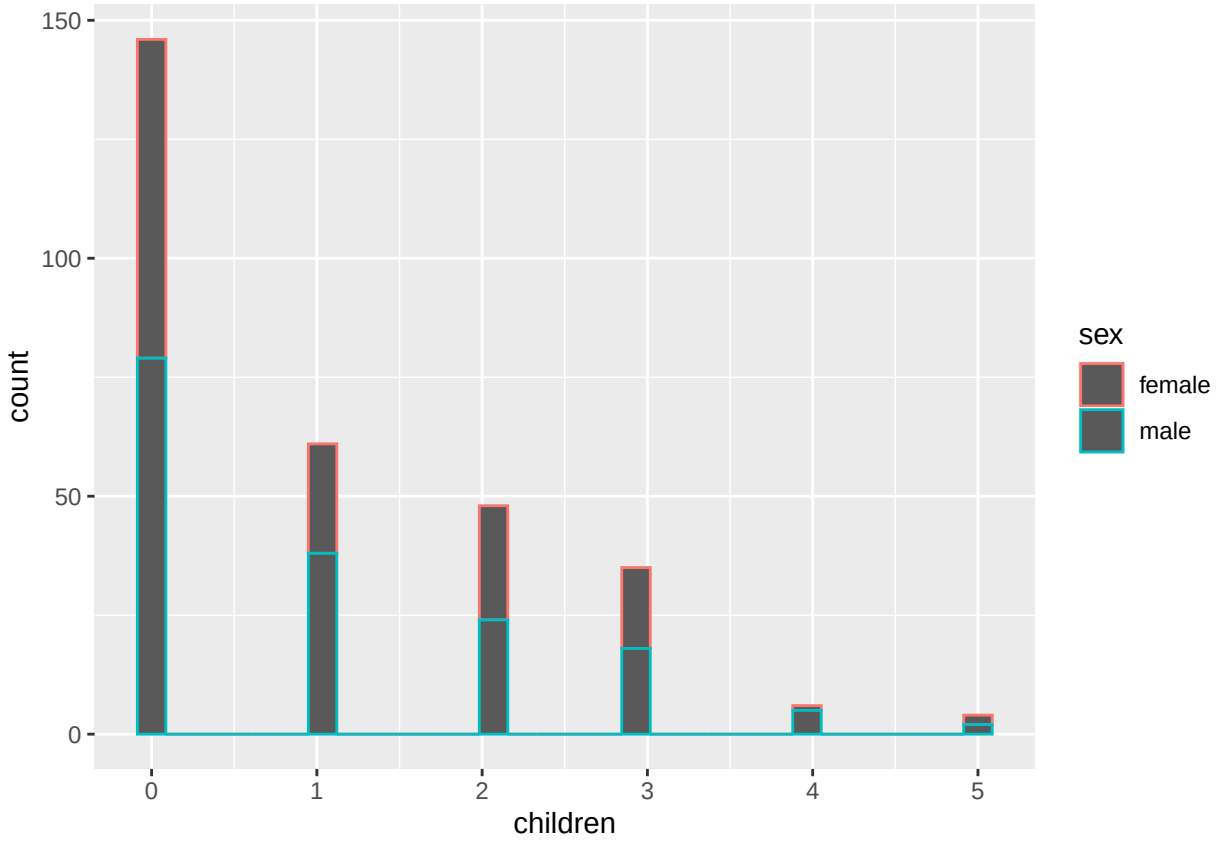

```
# box plot
# comparing that does region of living affects the smokers or not. Now, looking at graph one cans say t

ggplot(data = insurance_new ,
       aes(x=region, y=charges,shape=smoker,fill = smoker)) +
  geom_boxplot(color="black")+
  geom_smooth(method=lm)+
  theme_classic()+
  theme(legend.position="bottom")+
  geom_jitter(shape=16, position=position_jitter(0.2))
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# How many children per male/female.
ggplot(data = insurance_new , aes(x=children,color=sex)) +
 geom_histogram()
```

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



???? need to see what can be done.

```
insurance_new%>%
  filter(sex == "female")%>%
  count(sex,children,smoker,region)%>%
  arrange(sex, smoker)
```

```
## # A tibble: 29 x 5
##    sex    children smoker region        n
##    <fct>     <dbl> <fct>  <chr>     <int>
##  1 female        0 no     northeast    14
##  2 female        0 no     northwest    17
##  3 female        0 no     southeast    14
##  4 female        0 no     southwest    13
##  5 female        1 no     northeast     5
##  6 female        1 no     northwest     4
##  7 female        1 no     southeast     4
##  8 female        1 no     southwest     6
##  9 female        2 no     northeast     3
## 10 female        2 no     northwest     6
## # ... with 19 more rows
```

```
insurance_new
```

```
## # A tibble: 300 x 7
##      age sex      bmi children smoker region    charges
##    <dbl> <fct>  <dbl>    <dbl> <fct>  <chr>       <dbl>
##  1    63 female  25.1        0 no     northwest  14255.
##  2    18 male    38.2        0 yes    southeast  36308.
##  3    48 male    29.6        0 no     southwest  21232.
##  4    46 female  33.4        1 no     southeast   8241.
##  5    52 male    30.2        1 no     southwest   9725.
##  6    36 female  19.9        0 no     northeast   5458.
##  7    19 male    20.9        1 no     southwest   1832.
##  8    48 male    36.7        1 no     northwest  28469.
##  9    19 female  29.8        0 no     southwest   1744.
## 10    19 female  20.6        0 no     southwest   1732.
## # ... with 290 more rows
```

```
ab <- insurance_new%>%
  dplyr::select(age,smoker,sex)
ab
```

```
## # A tibble: 300 x 3
##      age smoker sex
##    <dbl> <fct>  <fct>
##  1    63 no     female
##  2    18 yes    male
##  3    48 no     male
##  4    46 no     female
##  5    52 no     male
##  6    36 no     female
##  7    19 no     male
##  8    48 no     male
##  9    19 no     female
## 10    19 no     female
## # ... with 290 more rows
```

2

2.1 multicollinearity plot

The response variable is not dependent on explanatory variable interns of multicollinearity.

The highest correlation is between charges and age with only 0.24. But if we exclude charges since charges is dependent variable, the highest correlation among the independent variables.

The age with bmi with only 0.04 which is nearly 0. So there exists no colinearity among the independent variables. This suggests that each of the variables might be useful if they are included in the regression model as they dont have any correlation with each other.

```
numeric_insurance <- cor(insurance_new[,c("bmi", "children", "age", "charges")])

numeric_insurance
```

```
##                  bmi    children        age   charges
## bmi       1.00000000 -0.01371482 0.04733455 0.1785191
## children -0.01371482  1.00000000 0.03529611 0.0781793
## age       0.04733455  0.03529611 1.00000000 0.2461625
## charges   0.17851911  0.07817930 0.24616249 1.0000000
```

we can also the scatter plots between the independent variables clearly there is no pattern that we can see verifying our output from the correlation matrix.

One can say from the graph that the points are independently plotted and one cannot find any kind of pattern on left side of graph. On the other hand one can identify the

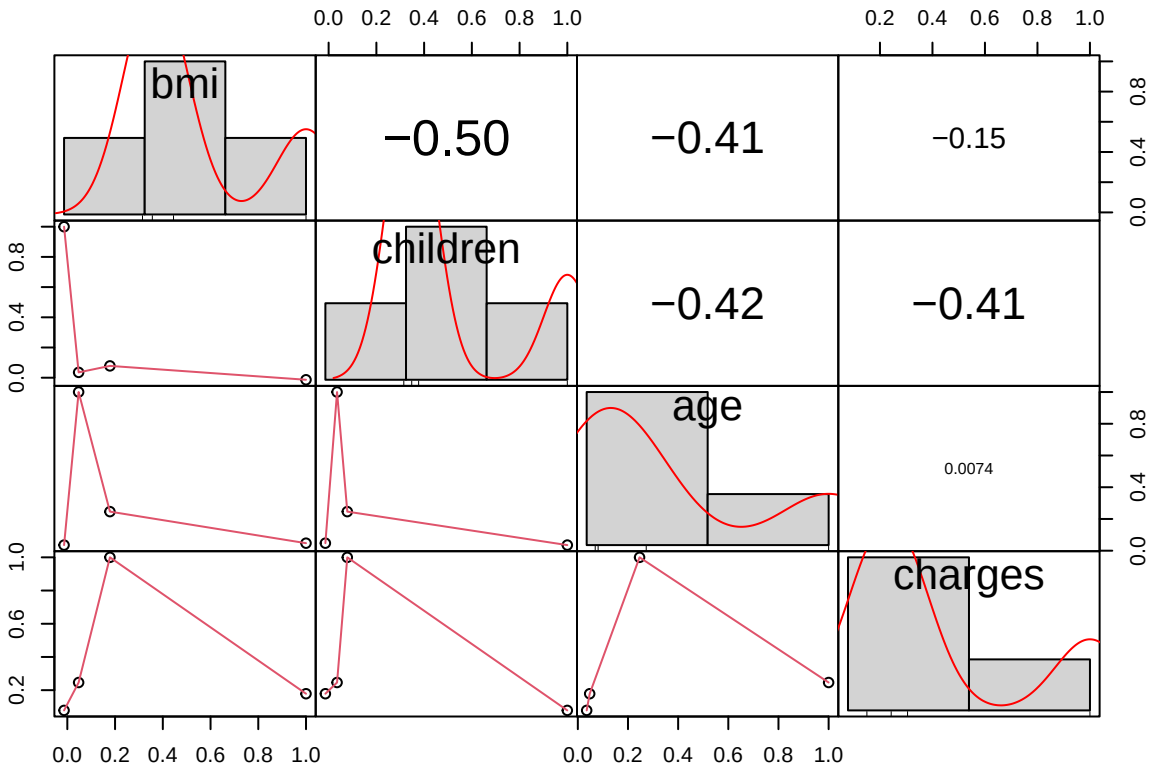


Figure 1: multicollinearity plot

2.2 Normality plot

```
# Light tailed at the end
qqnorm(insurance_new$bmi)
```

```
# right skewed
qqnorm(insurance_new$charges)
```

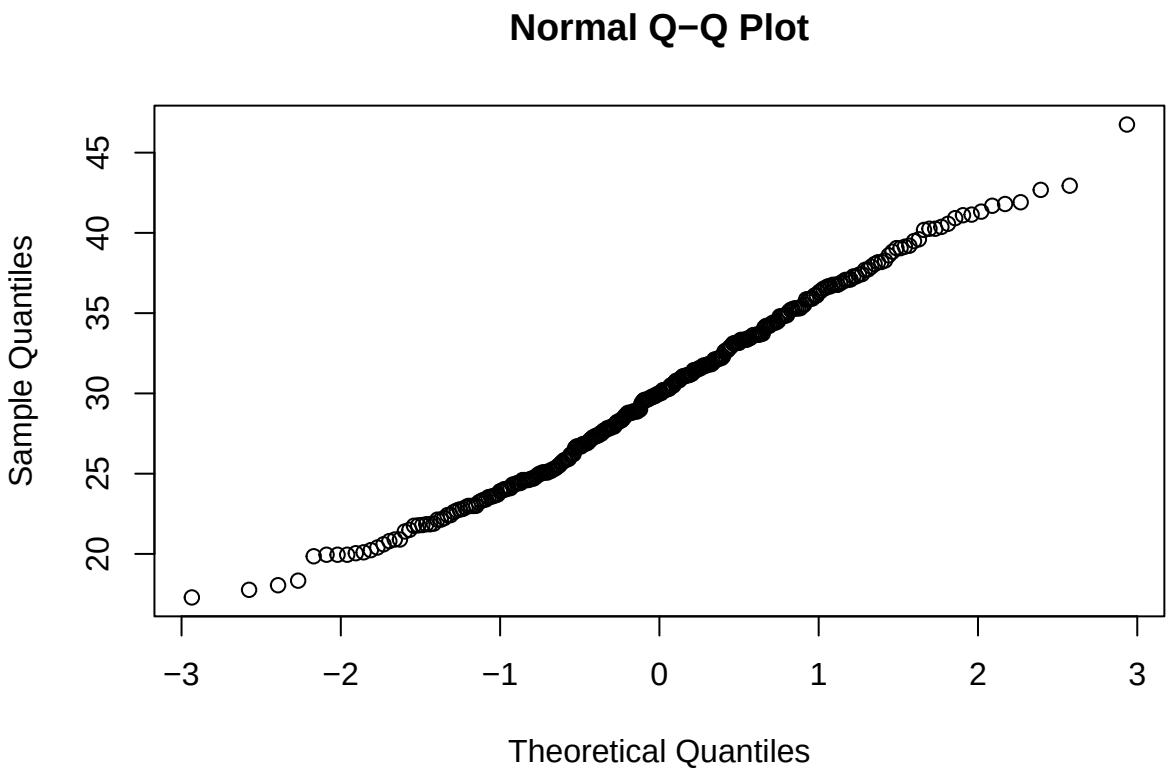


Figure 2: Q-Q plot

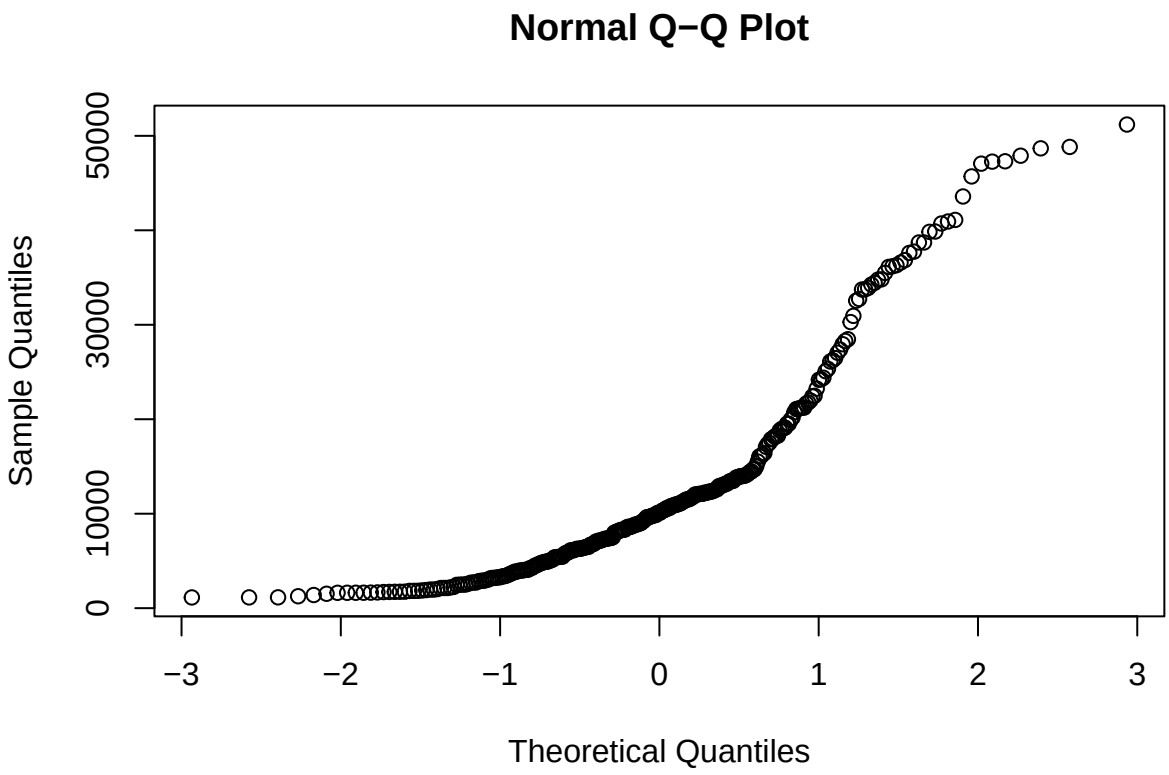


Figure 3: Q-Q plot

```
# Heavy tailed at the end
qqnorm(insurance_new$age)
```

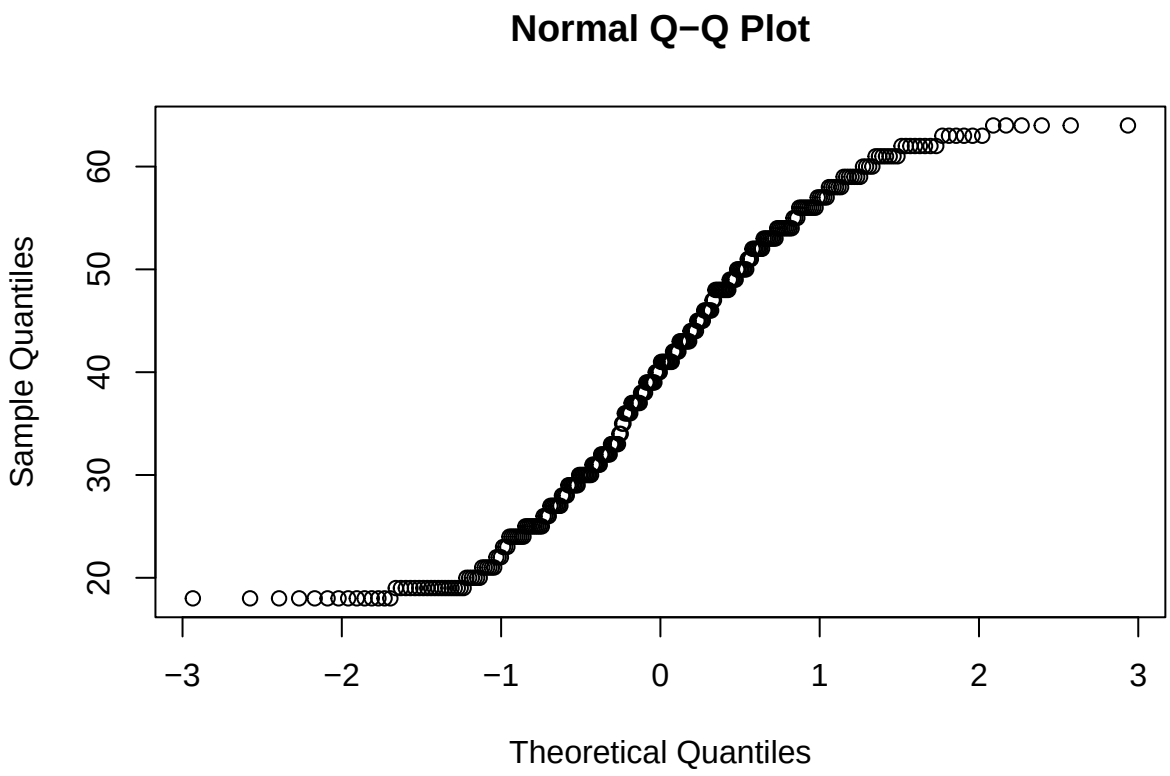


Figure 4: Q-Q plot

3

3.1 full regression model

Make use of the R generated dummy variable matrices ??

full regression model including both categorical and quantitative variables

```
lm(charges ~ age + children + bmi + region + sex + smoker,data = insurance_new) -> x
x
```

```
##
## Call:
## lm(formula = charges ~ age + children + bmi + region + sex +
##     smoker, data = insurance_new)
##
## Coefficients:
```

```
##     (Intercept)             age        children             bmi
##        -12033.2           261.1           532.8           353.3
## regionnorthwest regionsoutheast regionsouthwest         sexmale
##         -1545.5         -1505.4         -1719.9           607.8
##       smokeryes
##         22876.4
```

4

4.1 Fitted & residual values using matrix for quantitative variables

```
Xm <-  model.matrix(~age + children + bmi   , data=insurance_new )
Xm
```

```
##     (Intercept) age children    bmi
## 1             1  63        0 25.080
## 2             1  18        0 38.170
## 3             1  48        0 29.600
## 4             1  46        1 33.440
## 5             1  52        1 30.200
## 6             1  36        0 19.855
## 7             1  19        1 20.900
## 8             1  48        1 36.670
## 9             1  19        0 29.800
## 10            1  19        0 20.600
## 11            1  52        0 31.200
## 12            1  29        2 29.735
## 13            1  58        0 22.770
## 14            1  57        1 20.100
## 15            1  23        0 34.400
## 16            1  64        0 34.500
## 17            1  29        2 24.600
## 18            1  40        0 41.690
## 19            1  43        5 25.520
## 20            1  20        0 31.460
## 21            1  18        0 22.990
## 22            1  48        3 25.850
## 23            1  42        1 29.000
## 24            1  38        2 27.835
## 25            1  30        3 31.570
## 26            1  30        1 38.830
## 27            1  58        0 28.215
## 28            1  56        0 33.725
## 29            1  54        0 24.035
## 30            1  19        0 32.110
## 31            1  56        4 33.660
## 32            1  37        2 29.500
## 33            1  54        1 32.300
## 34            1  36        1 35.200
## 35            1  50        2 23.540
## 36            1  25        1 20.800
```

```
## 37          1  39        3 34.100
## 38          1  53        0 20.900
## 39          1  27        2 33.155
## 40          1  24        0 31.065
## 41          1  21        3 20.235
## 42          1  59        1 36.520
## 43          1  48        0 36.575
## 44          1  62        0 30.020
## 45          1  31        3 31.065
## 46          1  63        0 35.090
## 47          1  62        0 21.400
## 48          1  30        0 35.300
## 49          1  38        1 19.950
## 50          1  43        2 32.600
## 51          1  39        1 21.850
## 52          1  27        0 32.670
## 53          1  20        0 27.930
## 54          1  45        2 22.895
## 55          1  60        3 33.110
## 56          1  44        3 21.850
## 57          1  51        2 36.670
## 58          1  54        3 30.800
## 59          1  49        2 42.680
## 60          1  45        3 27.500
## 61          1  21        2 21.890
## 62          1  49        0 35.860
## 63          1  30        3 28.690
## 64          1  39        1 28.300
## 65          1  56        0 19.950
## 66          1  19        0 21.755
## 67          1  46        1 33.345
## 68          1  21        0 36.860
## 69          1  35        2 35.860
## 70          1  36        0 31.500
## 71          1  41        1 34.210
## 72          1  54        3 24.605
## 73          1  43        3 27.360
## 74          1  48        2 40.565
## 75          1  18        1 29.370
## 76          1  55        0 32.775
## 77          1  43        2 26.700
## 78          1  55        1 21.500
## 79          1  33        1 27.100
## 80          1  56        0 25.300
## 81          1  20        0 21.800
## 82          1  26        0 40.185
## 83          1  57        0 22.230
## 84          1  53        1 36.100
## 85          1  45        0 35.300
## 86          1  33        0 22.705
## 87          1  37        2 29.830
## 88          1  19        0 22.610
## 89          1  27        3 30.300
## 90          1  50        1 37.070
```

```
## 91            1  27        3 20.045
## 92            1  63        0 36.765
## 93            1  19        0 34.400
## 94            1  52        2 41.800
## 95            1  19        0 34.800
## 96            1  25        0 30.200
## 97            1  45        2 23.560
## 98            1  37        2 23.370
## 99            1  52        2 36.765
## 100           1  22        3 34.800
## 101           1  41        0 40.260
## 102           1  59        1 36.765
## 103           1  43        2 30.685
## 104           1  32        3 33.155
## 105           1  30        2 22.990
## 106           1  61        3 36.100
## 107           1  42        0 24.640
## 108           1  52        5 46.750
## 109           1  56        0 39.600
## 110           1  33        0 26.695
## 111           1  19        0 27.700
## 112           1  24        0 23.210
## 113           1  36        3 28.880
## 114           1  29        1 27.940
## 115           1  49        0 30.300
## 116           1  19        0 31.920
## 117           1  38        0 37.730
## 118           1  56        0 25.650
## 119           1  35        1 38.600
## 120           1  49        1 30.780
## 121           1  29        2 32.110
## 122           1  44        2 22.135
## 123           1  19        0 25.175
## 124           1  24        2 28.500
## 125           1  55        0 26.980
## 126           1  34        0 30.800
## 127           1  33        2 32.900
## 128           1  21        0 31.100
## 129           1  53        1 24.795
## 130           1  24        0 33.345
## 131           1  62        2 30.495
## 132           1  56        0 28.785
## 133           1  61        0 23.655
## 134           1  40        0 25.080
## 135           1  44        1 34.320
## 136           1  19        5 28.600
## 137           1  25        2 33.330
## 138           1  25        2 24.985
## 139           1  26        1 30.000
## 140           1  22        0 26.840
## 141           1  18        0 37.290
## 142           1  50        2 32.110
## 143           1  58        0 23.300
## 144           1  58        0 33.440
```

```
## 145          1  50        3 30.970
## 146          1  52        0 37.400
## 147          1  57        0 31.160
## 148          1  25        0 26.220
## 149          1  29        1 28.975
## 150          1  63        2 32.200
## 151          1  39        1 26.220
## 152          1  59        1 27.500
## 153          1  34        2 22.420
## 154          1  39        3 22.800
## 155          1  32        1 33.630
## 156          1  21        0 22.135
## 157          1  18        0 41.140
## 158          1  31        0 30.875
## 159          1  20        0 33.330
## 160          1  33        3 42.940
## 161          1  54        2 28.880
## 162          1  23        3 31.730
## 163          1  26        2 34.200
## 164          1  31        1 25.935
## 165          1  58        0 27.170
## 166          1  62        0 29.920
## 167          1  43        0 26.885
## 168          1  28        0 17.290
## 169          1  30        3 24.400
## 170          1  53        0 24.320
## 171          1  46        2 40.375
## 172          1  32        2 29.800
## 173          1  19        1 24.600
## 174          1  41        1 31.635
## 175          1  37        0 29.640
## 176          1  23        0 34.865
## 177          1  30        3 39.050
## 178          1  63        0 37.700
## 179          1  62        0 39.200
## 180          1  25        5 23.900
## 181          1  27        0 33.660
## 182          1  24        0 25.270
## 183          1  59        0 24.700
## 184          1  41        1 32.200
## 185          1  57        1 27.940
## 186          1  19        0 30.250
## 187          1  59        0 28.785
## 188          1  50        2 25.365
## 189          1  62        0 31.730
## 190          1  40        3 35.300
## 191          1  49        1 25.840
## 192          1  32        1 30.030
## 193          1  42        1 26.180
## 194          1  25        4 26.695
## 195          1  47        4 28.215
## 196          1  53        2 35.900
## 197          1  18        0 25.080
## 198          1  54        0 31.240
```

```
## 199          1  52        0 27.360
## 200          1  60        0 40.920
## 201          1  19        0 35.530
## 202          1  18        1 28.310
## 203          1  62        0 32.680
## 204          1  41        3 33.155
## 205          1  46        2 19.950
## 206          1  61        4 33.330
## 207          1  33        0 35.245
## 208          1  19        0 33.110
## 209          1  27        0 24.100
## 210          1  19        0 37.430
## 211          1  30        1 28.380
## 212          1  31        2 36.300
## 213          1  44        0 27.645
## 214          1  20        1 26.840
## 215          1  41        1 21.780
## 216          1  56        1 26.600
## 217          1  22        0 28.880
## 218          1  28        0 38.060
## 219          1  39        3 24.890
## 220          1  18        0 39.140
## 221          1  64        0 32.965
## 222          1  61        0 35.910
## 223          1  27        1 31.130
## 224          1  63        0 31.445
## 225          1  62        0 25.000
## 226          1  29        1 29.590
## 227          1  36        0 34.430
## 228          1  41        2 34.200
## 229          1  38        2 27.835
## 230          1  64        2 31.825
## 231          1  53        0 30.495
## 232          1  21        0 34.870
## 233          1  24        0 39.490
## 234          1  59        1 37.100
## 235          1  32        0 24.600
## 236          1  19        0 24.700
## 237          1  19        1 31.825
## 238          1  18        0 25.175
## 239          1  40        1 27.400
## 240          1  51        0 18.050
## 241          1  58        0 41.910
## 242          1  57        0 23.700
## 243          1  43        0 25.080
## 244          1  18        0 31.730
## 245          1  51        4 24.415
## 246          1  46        1 33.725
## 247          1  64        0 22.990
## 248          1  32        2 17.765
## 249          1  53        0 28.880
## 250          1  53        3 28.100
## 251          1  29        0 25.900
## 252          1  54        1 33.630
```

```
## 253           1  54        3 23.000
## 254           1  18        0 33.660
## 255           1  59        3 27.830
## 256           1  30        0 25.460
## 257           1  18        0 40.260
## 258           1  48        0 31.130
## 259           1  47        1 24.100
## 260           1  46        3 30.495
## 261           1  37        2 24.320
## 262           1  64        3 39.050
## 263           1  32        0 41.100
## 264           1  48        0 24.420
## 265           1  25        0 41.325
## 266           1  42        0 26.900
## 267           1  31        2 23.600
## 268           1  50        0 32.205
## 269           1  26        0 28.785
## 270           1  18        0 38.280
## 271           1  24        0 27.720
## 272           1  43        2 35.310
## 273           1  24        0 23.400
## 274           1  45        0 31.790
## 275           1  18        3 27.280
## 276           1  42        0 37.900
## 277           1  56        1 26.695
## 278           1  25        0 25.740
## 279           1  50        0 31.825
## 280           1  25        3 29.700
## 281           1  60        0 24.035
## 282           1  54        0 30.210
## 283           1  64        0 38.190
## 284           1  61        0 33.535
## 285           1  31        0 20.400
## 286           1  37        2 34.800
## 287           1  20        0 22.420
## 288           1  61        0 31.570
## 289           1  41        1 28.800
## 290           1  44        1 36.955
## 291           1  60        0 18.335
## 292           1  61        3 29.920
## 293           1  48        1 27.265
## 294           1  48        1 31.445
## 295           1  37        1 37.070
## 296           1  28        0 35.435
## 297           1  41        1 23.940
## 298           1  48        2 37.290
## 299           1  28        1 36.400
## 300           1  39        4 29.600
## attr(,"assign")
## [1] 0 1 2 3
```

```
Ym <- as.matrix(insurance_new%>%dplyr::select(charges))
Ym
```

```
##          charges
##   [1,] 14254.608
##   [2,] 36307.798
##   [3,] 21232.182
##   [4,]  8240.590
##   [5,]  9724.530
##   [6,]  5458.046
##   [7,]  1832.094
##   [8,] 28468.919
##   [9,]  1744.465
##  [10,]  1731.677
##  [11,]  9625.920
##  [12,] 18157.876
##  [13,] 11833.782
##  [14,] 12032.326
##  [15,]  1826.843
##  [16,] 13822.803
##  [17,]  4529.477
##  [18,]  5438.749
##  [19,] 14478.330
##  [20,]  1877.929
##  [21,]  1704.568
##  [22,] 24180.933
##  [23,]  7050.642
##  [24,]  7144.863
##  [25,]  4837.582
##  [26,] 18963.172
##  [27,] 12224.351
##  [28,] 10976.246
##  [29,] 10422.917
##  [30,]  2130.676
##  [31,] 12949.155
##  [32,]  6311.952
##  [33,] 11512.405
##  [34,] 38709.176
##  [35,] 10107.221
##  [36,]  3208.787
##  [37,]  7418.522
##  [38,] 21195.818
##  [39,]  4058.712
##  [40,] 34254.053
##  [41,]  3861.210
##  [42,] 28287.898
##  [43,]  8671.191
##  [44,] 13352.100
##  [45,]  5425.023
##  [46,] 47055.532
##  [47,] 12957.118
##  [48,] 36837.467
##  [49,]  5855.903
##  [50,]  7441.501
##  [51,]  6117.494
##  [52,]  2497.038
##  [53,]  1967.023
```

```
##  [54,] 21098.554
##  [55,] 13919.823
##  [56,]  8891.139
##  [57,] 10848.134
##  [58,] 12105.320
##  [59,]  9800.888
##  [60,]  8615.300
##  [61,]  3180.510
##  [62,]  8124.408
##  [63,] 20745.989
##  [64,] 21082.160
##  [65,] 22412.648
##  [66,]  1627.282
##  [67,]  8334.458
##  [68,]  1917.318
##  [69,]  5836.520
##  [70,]  4402.233
##  [71,]  6289.755
##  [72,] 12479.709
##  [73,]  8606.217
##  [74,] 45702.022
##  [75,]  1719.436
##  [76,] 10601.632
##  [77,] 22478.600
##  [78,] 10791.960
##  [79,] 19040.876
##  [80,] 11070.535
##  [81,] 20167.336
##  [82,]  3201.245
##  [83,] 12029.287
##  [84,] 10085.846
##  [85,]  7348.142
##  [86,] 21984.471
##  [87,]  6406.411
##  [88,]  1628.471
##  [89,]  4260.744
##  [90,]  9048.027
##  [91,] 16420.495
##  [92,] 13981.850
##  [93,]  1261.859
##  [94,] 47269.854
##  [95,] 34779.615
##  [96,] 33900.653
##  [97,]  8603.823
##  [98,]  6686.431
##  [99,] 26467.097
## [100,]  3443.064
## [101,]  5709.164
## [102,] 47896.791
## [103,]  8310.839
## [104,]  6128.797
## [105,] 17361.766
## [106,] 27941.288
## [107,] 19515.542
```

```
## [108,] 12592.534
## [109,] 10601.412
## [110,]  4571.413
## [111,] 16297.846
## [112,] 25081.768
## [113,]  6748.591
## [114,] 19107.780
## [115,]  8116.680
## [116,] 33750.292
## [117,]  5397.617
## [118,] 11454.022
## [119,]  4762.329
## [120,]  9778.347
## [121,]  4922.916
## [122,]  8302.536
## [123,]  1632.036
## [124,]  3537.703
## [125,] 11082.577
## [126,] 35491.640
## [127,]  5375.038
## [128,]  1526.312
## [129,] 10942.132
## [130,]  2855.438
## [131,] 15019.760
## [132,] 11658.379
## [133,] 13129.603
## [134,]  5415.661
## [135,]  7147.473
## [136,]  4687.797
## [137,] 36124.574
## [138,] 23241.475
## [139,]  2904.088
## [140,]  1665.000
## [141,]  1141.445
## [142,] 25333.333
## [143,] 11345.519
## [144,] 12231.614
## [145,] 10600.548
## [146,]  9634.538
## [147,] 43578.939
## [148,]  2721.321
## [149,]  4040.558
## [150,] 47305.305
## [151,]  6123.569
## [152,] 12333.828
## [153,] 27375.905
## [154,]  7985.815
## [155,] 37607.528
## [156,]  2585.851
## [157,]  1146.797
## [158,]  3857.759
## [159,]  1391.529
## [160,]  6360.994
## [161,] 12096.651
```

```
## [162,] 36189.102
## [163,]  3987.926
## [164,]  4239.893
## [165,] 12222.898
## [166,] 13457.961
## [167,] 21774.322
## [168,]  3732.625
## [169,] 18259.216
## [170,]  9863.472
## [171,]  8733.229
## [172,]  5152.134
## [173,]  1837.237
## [174,]  7358.176
## [175,]  5028.147
## [176,]  2899.489
## [177,] 40932.429
## [178,] 48824.450
## [179,] 13470.860
## [180,]  5080.096
## [181,]  2498.414
## [182,]  3044.213
## [183,] 12323.936
## [184,]  6775.961
## [185,] 11554.224
## [186,] 32548.340
## [187,] 12129.614
## [188,] 30284.643
## [189,] 14043.477
## [190,]  7196.867
## [191,]  9282.481
## [192,]  4074.454
## [193,]  7046.722
## [194,]  4877.981
## [195,] 10407.086
## [196,] 11163.568
## [197,]  2196.473
## [198,] 10338.932
## [199,] 24393.622
## [200,] 48673.559
## [201,]  1646.430
## [202,] 11272.331
## [203,] 13844.797
## [204,]  8538.288
## [205,]  9193.838
## [206,] 36580.282
## [207,] 12404.879
## [208,] 34439.856
## [209,]  2974.126
## [210,]  2138.071
## [211,] 19521.968
## [212,] 38711.000
## [213,]  7421.195
## [214,] 17085.268
## [215,]  6272.477
```

```
## [216,] 12044.342
## [217,]  2250.835
## [218,]  2689.495
## [219,] 21659.930
## [220,] 12890.058
## [221,] 14692.669
## [222,] 13635.638
## [223,] 34806.468
## [224,] 13974.456
## [225,] 13451.122
## [226,]  3947.413
## [227,] 37742.576
## [228,]  7261.741
## [229,]  6455.863
## [230,] 16069.085
## [231,] 10072.055
## [232,]  2020.552
## [233,]  2480.979
## [234,] 12347.172
## [235,] 17496.306
## [236,]  1737.376
## [237,]  2719.280
## [238,] 15518.180
## [239,]  6496.886
## [240,]  9644.253
## [241,] 24227.337
## [242,] 10959.330
## [243,]  7325.048
## [244,] 33732.687
## [245,] 11520.100
## [246,]  8823.986
## [247,] 27037.914
## [248,] 32734.186
## [249,]  9869.810
## [250,] 11741.726
## [251,]  3353.284
## [252,] 10825.254
## [253,] 12094.478
## [254,]  1136.399
## [255,] 14001.287
## [256,]  3645.089
## [257,]  1634.573
## [258,]  8280.623
## [259,] 26236.580
## [260,] 40720.551
## [261,]  6198.752
## [262,] 16085.128
## [263,]  3989.841
## [264,] 21223.676
## [265,] 17878.901
## [266,]  5969.723
## [267,]  4931.647
## [268,]  8835.265
## [269,]  3385.399
```

```
## [270,] 14133.038
## [271,]  2464.619
## [272,] 18806.145
## [273,]  1969.614
## [274,] 17929.303
## [275,] 18223.451
## [276,]  6474.013
## [277,] 26109.329
## [278,]  2137.654
## [279,] 41097.162
## [280,] 19933.458
## [281,] 13012.209
## [282,] 10231.500
## [283,] 14410.932
## [284,] 13143.337
## [285,]  3260.199
## [286,] 39836.519
## [287,] 14711.744
## [288,] 12557.605
## [289,]  6282.235
## [290,]  8023.135
## [291,] 13204.286
## [292,] 30942.192
## [293,]  9447.250
## [294,]  8964.061
## [295,] 39871.704
## [296,]  3268.847
## [297,]  6858.480
## [298,]  8978.185
## [299,] 51194.559
## [300,]  7512.267
```

```
# Let's use R code to establish matrix X :
#-----------------------------------------------------------------------
#  A = (X^T*X)^-1*X^TY
(solve(t(Xm)%*%Xm))%*%(t(Xm)%*%Ym)
```

```
##                charges
## (Intercept) -4825.6927
## age           185.6491
## children      669.1986
## bmi           333.8682
```

```
#-----------------------------------------------------------------------
# fitted values
Xm%*%((solve(t(Xm)%*%Xm))%*%(t(Xm)%*%Ym)) -> fitted_values
#-----------------------------------------------------------------------
# residual values
Ym-Xm%*%((solve(t(Xm)%*%Xm))%*%(t(Xm)%*%Ym)) -> residual_values
#-----------------------------------------------------------------------
# producing the table of residula and fitted plot for the model.
matrix = data.frame(fitted_values, residual_values)
names(matrix)[1] <- "fitted_values"
```

```r
names(matrix)[2] <- "residual_values"
matrix
```

```
##     fitted_values residual_values
## 1        15243.617       -989.0085
## 2        11259.742      25048.0563
## 3        13967.964       7264.2178
## 4        15547.919      -7307.3293
## 5        15580.081      -5855.5505
## 6         8486.629      -3028.5827
## 7         6348.685      -4516.5912
## 8        16997.612      11471.3074
## 9         8650.914      -6906.4489
## 10        5579.326      -3847.6491
## 11       15244.750      -5618.8301
## 12       11824.101       6333.7752
## 13       13544.136      -1710.3532
## 14       13136.257      -1103.9309
## 15       10929.304      -9102.4613
## 16       18574.305      -4751.5016
## 17       10109.687      -5580.2104
## 18       16519.239     -11080.4895
## 19       15023.530       -545.1995
## 20        9390.784      -7512.8549
## 21        6191.622      -4487.0540
## 22       14723.554       9457.3790
## 23       13322.948      -6272.3055
## 24       12860.593      -5715.7305
## 25       13291.597      -8454.0145
## 26       14377.083       4586.0890
## 27       15362.048      -3137.6972
## 28       16830.364      -5854.1181
## 29       13223.882      -2800.9658
## 30        9422.150      -7291.4736
## 31       19485.457      -6536.3015
## 32       13230.835      -6918.8827
## 33       16652.502      -5140.0970
## 34       14279.036      24430.1401
## 35       13654.418      -3547.1978
## 36        7429.193      -4220.4061
## 37       15807.125      -8388.6034
## 38       11991.556       9204.2616
## 39       12594.632      -8535.9195
## 40       10001.503      24252.5506
## 41        7836.358      -3975.1487
## 42       18989.672       9298.2261
## 43       16296.695      -7625.5042
## 44       16707.277      -3355.1769
## 45       13308.642      -7883.6191
## 46       18585.638      28469.8943
## 47       13829.332       -872.2145
## 48       12529.329      24308.1376
## 49        9558.844      -3702.9410
```

```
## 50      15379.721      -7938.2199
## 51      10378.842      -4261.3478
## 52      11094.309      -8597.2703
## 53       8212.229      -6245.2067
## 54      12510.828       8587.7262
## 55      19375.227      -5455.4043
## 56      12645.485      -3754.3456
## 57      18223.758      -7375.6232
## 58      17490.097      -5384.7769
## 59      19859.007     -10058.1192
## 60      14717.490      -6102.1897
## 61       7719.712      -4539.2015
## 62      16243.629      -8119.2204
## 63      12330.056       8415.9328
## 64      12532.292       8549.8676
## 65      12231.329      10181.3196
## 66       5964.944      -4337.6615
## 67      15516.201      -7181.7439
## 68      11379.322      -9462.0035
## 69      14982.938      -9146.4181
## 70      12374.525      -7972.2918
## 71      14876.752      -8586.9970
## 72      15421.783      -2942.0742
## 73      14299.450      -5693.2326
## 74      18967.227      26734.7954
## 75       8990.900      -7271.4638
## 76      16327.540      -5725.9077
## 77      13409.898       9068.7017
## 78      13232.374      -2440.4142
## 79      11017.756       8023.1201
## 80      14017.524      -2946.9890
## 81       6165.617      14001.7189
## 82      13417.679     -10216.4342
## 83      13178.198      -1148.9109
## 84      17735.552      -7649.7062
## 85      15314.066      -7965.9241
## 86       8881.206      13103.2643
## 87      13341.011      -6934.6005
## 88       6250.401      -4621.9304
## 89      12310.637      -8049.8928
## 90      17502.457      -8454.4298
## 91       8886.818       7533.6765
## 92      19144.867      -5163.0167
## 93      10186.708      -8924.8488
## 94      20122.151      27147.7033
## 95      10320.255      24459.3599
## 96       9898.356      24002.2971
## 97      12732.850      -4129.0268
## 98      11184.222      -4497.7911
## 99      18441.124       8025.9733
## 100     12884.798      -9441.7343
## 101     16227.456     -10518.2917
## 102     19071.469      28825.3221
## 103     14740.363      -6429.5241
```

```
## 104     14192.076      -8063.2787
## 105      9757.809       7603.9574
## 106     20559.142       7382.1452
## 107     11198.083       8317.4582
## 108     23782.394     -11189.8598
## 109     18791.840      -8190.4278
## 110     10213.341      -5641.9275
## 111      7949.791       8348.0554
## 112      7378.968      17702.8001
## 113     13507.386      -6758.7947
## 114     10555.609       8552.1709
## 115     14387.321      -6270.6414
## 116      9358.715      24391.5772
## 117     14825.822      -9428.2055
## 118     14134.378      -2680.3563
## 119     15228.539     -10466.2098
## 120     15216.777      -5438.4295
## 121     12617.038      -7694.1220
## 122     12071.439      -3768.9032
## 123      7106.773      -5474.7371
## 124     10483.528      -6945.8250
## 125     14392.773      -3310.1963
## 126     11769.519      23722.1212
## 127     13623.390      -8248.3523
## 128      9456.241      -7929.9288
## 129     13961.172      -3019.0397
## 130     10762.722      -7907.2848
## 131     18204.261      -3184.5013
## 132     15181.055      -3522.6756
## 133     14396.556      -1266.9528
## 134     10973.687      -5558.0260
## 135     15470.425      -8322.9520
## 136     11596.265      -6908.4682
## 137     12281.761      23842.8130
## 138      9495.630      13745.8443
## 139     10686.430      -7782.3420
## 140      8219.611      -6554.6117
## 141     10965.938      -9824.4928
## 142     16515.669       8817.6636
## 143     13721.086      -2375.5667
## 144     17106.510      -4874.8960
## 145     16804.258      -6203.7098
## 146     17314.733      -7680.1952
## 147     16159.641      27419.2985
## 148      8569.560      -5848.2395
## 149     10901.162      -6860.6041
## 150     18959.156      28346.1492
## 151     11837.846      -5714.2777
## 152     15978.180      -3644.3520
## 153     10310.100      17065.8045
## 154     12034.414      -4048.5993
## 155     13012.266      24595.2614
## 156      6463.112      -3877.2614
## 157     12251.331     -11104.5340
```

```
## 158     11237.612      -7379.8524
## 159     10015.118      -8623.5892
## 160     17644.626     -11283.6324
## 161     16179.871      -4083.2201
## 162     12045.472      24143.6298
## 163     12757.875      -8769.9492
## 164     10257.501      -6017.6085
## 165     15013.156      -2790.2575
## 166     16673.890      -3215.9291
## 167     12133.267       9641.0555
## 168      6145.064      -2412.4392
## 169     10897.762       7361.4545
## 170     13133.386      -3269.9139
## 171     18532.494      -9799.2645
## 172     12402.750      -7250.6156
## 173      7583.998      -5746.7607
## 174     14017.041      -6658.8656
## 175     11939.179      -6911.0324
## 176     11084.553      -8185.0636
## 177     15788.931      25143.4983
## 178     19457.034      29367.4161
## 179     19772.187      -6301.3271
## 180     11140.979      -6060.8831
## 181     11424.838      -8926.4238
## 182      8066.736      -5022.5230
## 183     14374.150      -2050.2143
## 184     14205.677      -7429.7158
## 185     15753.784      -4199.5602
## 186      8801.155      23747.1859
## 187     15738.002      -3608.3879
## 188     14263.728      16020.9150
## 189     17278.191      -3234.7147
## 190     16393.416      -9196.5494
## 191     13567.468      -4284.9871
## 192     11810.341      -7735.8870
## 193     12381.439      -5334.7169
## 194     11404.942      -6526.9612
## 195     15996.702      -5589.6165
## 196     18337.977      -7174.4092
## 197      6889.407      -4692.9335
## 198     15629.403      -5290.4715
## 199     13962.696      10430.9263
## 200     19975.142      28698.4165
## 201     10563.979      -8917.5492
## 202      8637.000       2635.3316
## 203     17595.366      -3750.5690
## 204     15862.918      -7324.6297
## 205     11713.235      -2519.3965
## 206     20303.526      16276.7562
## 207     13067.914       -663.0349
## 208      9756.018      24683.8381
## 209      8233.058      -5258.9318
## 210     11198.329      -9060.2579
## 211     10888.160       8633.8083
```

```
## 212     14387.244      24323.7560
## 213     12572.656      -5151.4611
## 214      8517.512       8567.7559
## 215     10726.770      -4454.2925
## 216     15120.751      -3076.4093
## 217      8900.702      -6649.8673
## 218     13079.508     -10390.0122
## 219     12732.199       8927.7311
## 220     11583.594       1306.4635
## 221     18061.817      -3369.1475
## 222     18488.112      -4852.4736
## 223     11249.350      23557.1175
## 224     17368.688      -3394.2325
## 225     15031.258      -1580.1362
## 226     11106.491      -7159.0782
## 227     13352.759      24389.8170
## 228     15542.612      -8280.8709
## 229     12860.593      -6404.7305
## 230     19019.604      -2950.5196
## 231     15195.022      -5122.9671
## 232     10714.924      -8694.3718
## 233     12814.343     -10333.3636
## 234     19183.315      -6836.1431
## 235      9328.238       8168.0685
## 236      6948.186      -5210.8099
## 237      9996.196      -7276.9160
## 238      6921.124       8597.0561
## 239     12417.460      -5920.5741
## 240     10668.734      -1024.4812
## 241     19934.374       4292.9636
## 242     13668.984      -2709.6539
## 243     11530.634      -4205.5863
## 244      9109.631      24623.0562
## 245     15470.600      -3950.4997
## 246     15643.071      -6819.0856
## 247     14731.481      12306.4329
## 248      8384.645      24349.5409
## 249     14655.825      -4786.0147
## 250     16403.004      -4661.2776
## 251      9205.319      -5852.0349
## 252     17096.547      -6271.2931
## 253     14885.925      -2791.4467
## 254      9753.996      -8617.5968
## 255     17426.754      -3425.4671
## 256      9244.066      -5598.9766
## 257     11957.527     -10322.9532
## 258     14478.783      -6198.1602
## 259     12615.239      13621.3413
## 260     15903.074      24817.4768
## 261     11501.397      -5302.6454
## 262     22101.001      -6015.8735
## 263     14837.063     -10847.2225
## 264     12238.527       8985.1488
## 265     13612.640       4266.2607
```

```
## 266     11952.626      -5982.9026
## 267     10147.117      -5215.4704
## 268     15208.989      -6373.7245
## 269      9611.581      -6226.1823
## 270     11296.467       2836.5703
## 271      8884.714      -6420.0947
## 272     16284.504       2521.6416
## 273      7442.403      -5472.7887
## 274     14142.189       3787.1148
## 275      9631.513       8591.9385
## 276     15625.176      -9151.1632
## 277     15152.469      10956.8603
## 278      8409.304      -6271.6499
## 279     15082.120      26015.0422
## 280     11739.018       8194.4404
## 281     14337.777      -1325.5684
## 282     15285.519      -5054.0189
## 283     19806.278      -5395.3463
## 284     17695.174      -4551.8378
## 285      7740.342      -4480.1428
## 286     15000.336      24836.1826
## 287      6372.615       8339.1284
## 288     17039.123      -4481.5181
## 289     13070.525      -6788.2897
## 290     16350.168      -8327.0321
## 291     12434.728        769.5575
## 292     18495.837      12446.3552
## 293     13857.581      -4410.3304
## 294     15253.150      -6289.0895
## 295     15089.019      24782.6857
## 296     12203.103      -8934.2568
## 297     11447.925      -4589.4455
## 298     17873.808      -8895.6234
## 299     13194.485      38000.0742
## 300     14973.917      -7461.6500
```

```
#-----------------------------------------------------------------------
# Both fitted values and residual values match with matrix model
lm(charges ~ age + children + bmi, data= insurance_new) -> AB
summary(AB)
```

```
##
## Call:
## lm(formula = charges ~ age + children + bmi, data = insurance_new)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -11284  -6913  -4546   5023  38000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4825.69    3744.43  -1.289  0.19849
## age           185.65      43.67   4.251 2.85e-05 ***
## children      669.20     513.46   1.303  0.19348
```

```
## bmi           333.87     109.87   3.039  0.00259 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10910 on 296 degrees of freedom
## Multiple R-squared:  0.0937, Adjusted R-squared:  0.08452
## F-statistic:  10.2 on 3 and 296 DF,  p-value: 2.061e-06
```

```
residuals(AB)
```

```
##          1          2          3          4          5          6
##   -989.0085  25048.0563   7264.2178  -7307.3293  -5855.5505  -3028.5827
##          7          8          9         10         11         12
##  -4516.5912  11471.3074  -6906.4489  -3847.6491  -5618.8301   6333.7752
##         13         14         15         16         17         18
##  -1710.3532  -1103.9309  -9102.4613  -4751.5016  -5580.2104 -11080.4895
##         19         20         21         22         23         24
##   -545.1995  -7512.8549  -4487.0540   9457.3790  -6272.3055  -5715.7305
##         25         26         27         28         29         30
##  -8454.0145   4586.0890  -3137.6972  -5854.1181  -2800.9658  -7291.4736
##         31         32         33         34         35         36
##  -6536.3015  -6918.8827  -5140.0970  24430.1401  -3547.1978  -4220.4061
##         37         38         39         40         41         42
##  -8388.6034   9204.2616  -8535.9195  24252.5506  -3975.1487   9298.2261
##         43         44         45         46         47         48
##  -7625.5042  -3355.1769  -7883.6191  28469.8943   -872.2145  24308.1376
##         49         50         51         52         53         54
##  -3702.9410  -7938.2199  -4261.3478  -8597.2703  -6245.2067   8587.7262
##         55         56         57         58         59         60
##  -5455.4043  -3754.3456  -7375.6232  -5384.7769 -10058.1192  -6102.1897
##         61         62         63         64         65         66
##  -4539.2015  -8119.2204   8415.9328   8549.8676  10181.3196  -4337.6615
##         67         68         69         70         71         72
##  -7181.7439  -9462.0035  -9146.4181  -7972.2918  -8586.9970  -2942.0742
##         73         74         75         76         77         78
##  -5693.2326  26734.7954  -7271.4638  -5725.9077   9068.7017  -2440.4142
##         79         80         81         82         83         84
##   8023.1201  -2946.9890  14001.7189 -10216.4342  -1148.9109  -7649.7062
##         85         86         87         88         89         90
##  -7965.9241  13103.2643  -6934.6005  -4621.9304  -8049.8928  -8454.4298
##         91         92         93         94         95         96
##   7533.6765  -5163.0167  -8924.8488  27147.7033  24459.3599  24002.2971
##         97         98         99        100        101        102
##  -4129.0268  -4497.7911   8025.9733  -9441.7343 -10518.2917  28825.3221
##        103        104        105        106        107        108
##  -6429.5241  -8063.2787   7603.9574   7382.1452   8317.4582 -11189.8598
##        109        110        111        112        113        114
##  -8190.4278  -5641.9275   8348.0554  17702.8001  -6758.7947   8552.1709
##        115        116        117        118        119        120
##  -6270.6414  24391.5772  -9428.2055  -2680.3563 -10466.2098  -5438.4295
##        121        122        123        124        125        126
##  -7694.1220  -3768.9032  -5474.7371  -6945.8250  -3310.1963  23722.1212
##        127        128        129        130        131        132
##  -8248.3523  -7929.9288  -3019.0397  -7907.2848  -3184.5013  -3522.6756
```

```
##         133         134         135         136         137         138
##  -1266.9528  -5558.0260  -8322.9520  -6908.4682  23842.8130  13745.8443
##         139         140         141         142         143         144
##  -7782.3420  -6554.6117  -9824.4928   8817.6636  -2375.5667  -4874.8960
##         145         146         147         148         149         150
##  -6203.7098  -7680.1952  27419.2985  -5848.2395  -6860.6041  28346.1492
##         151         152         153         154         155         156
##  -5714.2777  -3644.3520  17065.8045  -4048.5993  24595.2614  -3877.2614
##         157         158         159         160         161         162
## -11104.5340  -7379.8524  -8623.5892 -11283.6324  -4083.2201  24143.6298
##         163         164         165         166         167         168
##  -8769.9492  -6017.6085  -2790.2575  -3215.9291   9641.0555  -2412.4392
##         169         170         171         172         173         174
##   7361.4545  -3269.9139  -9799.2645  -7250.6156  -5746.7607  -6658.8656
##         175         176         177         178         179         180
##  -6911.0324  -8185.0636  25143.4983  29367.4161  -6301.3271  -6060.8831
##         181         182         183         184         185         186
##  -8926.4238  -5022.5230  -2050.2143  -7429.7158  -4199.5602  23747.1859
##         187         188         189         190         191         192
##  -3608.3879  16020.9150  -3234.7147  -9196.5494  -4284.9871  -7735.8870
##         193         194         195         196         197         198
##  -5334.7169  -6526.9612  -5589.6165  -7174.4092  -4692.9335  -5290.4715
##         199         200         201         202         203         204
##  10430.9263  28698.4165  -8917.5492   2635.3316  -3750.5690  -7324.6297
##         205         206         207         208         209         210
##  -2519.3965  16276.7562   -663.0349  24683.8381  -5258.9318  -9060.2579
##         211         212         213         214         215         216
##   8633.8083  24323.7560  -5151.4611   8567.7559  -4454.2925  -3076.4093
##         217         218         219         220         221         222
##  -6649.8673 -10390.0122   8927.7311   1306.4635  -3369.1475  -4852.4736
##         223         224         225         226         227         228
##  23557.1175  -3394.2325  -1580.1362  -7159.0782  24389.8170  -8280.8709
##         229         230         231         232         233         234
##  -6404.7305  -2950.5196  -5122.9671  -8694.3718 -10333.3636  -6836.1431
##         235         236         237         238         239         240
##   8168.0685  -5210.8099  -7276.9160   8597.0561  -5920.5741  -1024.4812
##         241         242         243         244         245         246
##   4292.9636  -2709.6539  -4205.5863  24623.0562  -3950.4997  -6819.0856
##         247         248         249         250         251         252
##  12306.4329  24349.5409  -4786.0147  -4661.2776  -5852.0349  -6271.2931
##         253         254         255         256         257         258
##  -2791.4467  -8617.5968  -3425.4671  -5598.9766 -10322.9532  -6198.1602
##         259         260         261         262         263         264
##  13621.3413  24817.4768  -5302.6454  -6015.8735 -10847.2225   8985.1488
##         265         266         267         268         269         270
##   4266.2607  -5982.9026  -5215.4704  -6373.7245  -6226.1823   2836.5703
##         271         272         273         274         275         276
##  -6420.0947   2521.6416  -5472.7887   3787.1148   8591.9385  -9151.1632
##         277         278         279         280         281         282
##  10956.8603  -6271.6499  26015.0422   8194.4404  -1325.5684  -5054.0189
##         283         284         285         286         287         288
##  -5395.3463  -4551.8378  -4480.1428  24836.1826   8339.1284  -4481.5181
##         289         290         291         292         293         294
##  -6788.2897  -8327.0321    769.5575  12446.3552  -4410.3304  -6289.0895
```

```
##         295         296         297         298         299         300
##  24782.6857  -8934.2568  -4589.4455  -8895.6234  38000.0742  -7461.6500
```

```
fitted(AB)
```

```
##         1         2         3         4         5         6         7         8
## 15243.617 11259.742 13967.964 15547.919 15580.081  8486.629  6348.685 16997.612
##         9        10        11        12        13        14        15        16
##  8650.914  5579.326 15244.750 11824.101 13544.136 13136.257 10929.304 18574.305
##        17        18        19        20        21        22        23        24
## 10109.687 16519.239 15023.530  9390.784  6191.622 14723.554 13322.948 12860.593
##        25        26        27        28        29        30        31        32
## 13291.597 14377.083 15362.048 16830.364 13223.882  9422.150 19485.457 13230.835
##        33        34        35        36        37        38        39        40
## 16652.502 14279.036 13654.418  7429.193 15807.125 11991.556 12594.632 10001.503
##        41        42        43        44        45        46        47        48
##  7836.358 18989.672 16296.695 16707.277 13308.642 18585.638 13829.332 12529.329
##        49        50        51        52        53        54        55        56
##  9558.844 15379.721 10378.842 11094.309  8212.229 12510.828 19375.227 12645.485
##        57        58        59        60        61        62        63        64
## 18223.758 17490.097 19859.007 14717.490  7719.712 16243.629 12330.056 12532.292
##        65        66        67        68        69        70        71        72
## 12231.329  5964.944 15516.201 11379.322 14982.938 12374.525 14876.752 15421.783
##        73        74        75        76        77        78        79        80
## 14299.450 18967.227  8990.900 16327.540 13409.898 13232.374 11017.756 14017.524
##        81        82        83        84        85        86        87        88
##  6165.617 13417.679 13178.198 17735.552 15314.066  8881.206 13341.011  6250.401
##        89        90        91        92        93        94        95        96
## 12310.637 17502.457  8886.818 19144.867 10186.708 20122.151 10320.255  9898.356
##        97        98        99       100       101       102       103       104
## 12732.850 11184.222 18441.124 12884.798 16227.456 19071.469 14740.363 14192.076
##       105       106       107       108       109       110       111       112
##  9757.809 20559.142 11198.083 23782.394 18791.840 10213.341  7949.791  7378.968
##       113       114       115       116       117       118       119       120
## 13507.386 10555.609 14387.321  9358.715 14825.822 14134.378 15228.539 15216.777
##       121       122       123       124       125       126       127       128
## 12617.038 12071.439  7106.773 10483.528 14392.773 11769.519 13623.390  9456.241
##       129       130       131       132       133       134       135       136
## 13961.172 10762.722 18204.261 15181.055 14396.556 10973.687 15470.425 11596.265
##       137       138       139       140       141       142       143       144
## 12281.761  9495.630 10686.430  8219.611 10965.938 16515.669 13721.086 17106.510
##       145       146       147       148       149       150       151       152
## 16804.258 17314.733 16159.641  8569.560 10901.162 18959.156 11837.846 15978.180
##       153       154       155       156       157       158       159       160
## 10310.100 12034.414 13012.266  6463.112 12251.331 11237.612 10015.118 17644.626
##       161       162       163       164       165       166       167       168
## 16179.871 12045.472 12757.875 10257.501 15013.156 16673.890 12133.267  6145.064
##       169       170       171       172       173       174       175       176
## 10897.762 13133.386 18532.494 12402.750  7583.998 14017.041 11939.179 11084.553
##       177       178       179       180       181       182       183       184
## 15788.931 19457.034 19772.187 11140.979 11424.838  8066.736 14374.150 14205.677
##       185       186       187       188       189       190       191       192
## 15753.784  8801.155 15738.002 14263.728 17278.191 16393.416 13567.468 11810.341
##       193       194       195       196       197       198       199       200
```

```
## 12381.439 11404.942 15996.702 18337.977  6889.407 15629.403 13962.696 19975.142
##       201       202       203       204       205       206       207       208
## 10563.979  8637.000 17595.366 15862.918 11713.235 20303.526 13067.914  9756.018
##       209       210       211       212       213       214       215       216
##  8233.058 11198.329 10888.160 14387.244 12572.656  8517.512 10726.770 15120.751
##       217       218       219       220       221       222       223       224
##  8900.702 13079.508 12732.199 11583.594 18061.817 18488.112 11249.350 17368.688
##       225       226       227       228       229       230       231       232
## 15031.258 11106.491 13352.759 15542.612 12860.593 19019.604 15195.022 10714.924
##       233       234       235       236       237       238       239       240
## 12814.343 19183.315  9328.238  6948.186  9996.196  6921.124 12417.460 10668.734
##       241       242       243       244       245       246       247       248
## 19934.374 13668.984 11530.634  9109.631 15470.600 15643.071 14731.481  8384.645
##       249       250       251       252       253       254       255       256
## 14655.825 16403.004  9205.319 17096.547 14885.925  9753.996 17426.754  9244.066
##       257       258       259       260       261       262       263       264
## 11957.527 14478.783 12615.239 15903.074 11501.397 22101.001 14837.063 12238.527
##       265       266       267       268       269       270       271       272
## 13612.640 11952.626 10147.117 15208.989  9611.581 11296.467  8884.714 16284.504
##       273       274       275       276       277       278       279       280
##  7442.403 14142.189  9631.513 15625.176 15152.469  8409.304 15082.120 11739.018
##       281       282       283       284       285       286       287       288
## 14337.777 15285.519 19806.278 17695.174  7740.342 15000.336  6372.615 17039.123
##       289       290       291       292       293       294       295       296
## 13070.525 16350.168 12434.728 18495.837 13857.581 15253.150 15089.019 12203.103
##       297       298       299       300
## 11447.925 17873.808 13194.485 14973.917
```

4.2 Matrix method with both quantitative and qualitative variables(dummy variables included automatically)

```
# The results are the same using both the matrix method and lm method.

Xm <-  model.matrix(~age + children + bmi + region + sex+ smoker  , data=insurance_new )
Ym <- as.matrix(insurance_new%>%dplyr::select(charges))

# Let's use R code to establish matrix X :
#-----------------------------------------------------------------------
#  A = (X^T*X)^-1*X^TY
(solve(t(Xm)%*%Xm))%*%(t(Xm)%*%Ym)
```

```
##                     charges
## (Intercept)     -12033.2491
## age               261.1443
## children          532.7552
## bmi               353.3493
## regionnorthwest -1545.5302
## regionsoutheast -1505.3521
## regionsouthwest -1719.9011
## sexmale           607.7807
## smokeryes       22876.3789
```

```r
#-----------------------------------------------------------------------
# fitted values
Xm%*%((solve(t(Xm)%*%Xm))%*%(t(Xm)%*%Ym)) -> fitted_values
#-----------------------------------------------------------------------
# residual values
Ym-Xm%*%((solve(t(Xm)%*%Xm))%*%(t(Xm)%*%Ym)) -> residual_values
#-----------------------------------------------------------------------
# producing the table of residula and fitted plot for the model.
matrix = data.frame(fitted_values, residual_values)
names(matrix)[1] <- "fitted_values"
names(matrix)[2] <- "residual_values"
matrix
```

```
##     fitted_values residual_values
## 1     11735.31061     2519.297588
## 2     28133.49680     8174.301498
## 3      9848.69459    11383.487674
## 4     10822.79067    -2582.201071
## 5     11638.03656    -1913.506557
## 6      4383.69497     1074.351480
## 7      -265.87323     2097.967229
## 8     13054.00001    15414.919004
## 9      1738.39921        6.065788
## 10    -1512.41394     3244.090940
## 11    10850.84988    -1224.929878
## 12     6174.53657    11983.339430
## 13     9653.53040     2180.251898
## 14     8767.14985     3265.176147
## 15     5016.16366    -3189.320658
## 16    15758.41465    -1935.611645
## 17     3577.93651      951.540493
## 18    12246.08172    -6807.332617
## 19     9979.63334     4498.696808
## 20     2800.65225     -922.722854
## 21     1398.62828      305.939817
## 22    32605.04783    -8424.114330
## 23     7994.79382     -944.151815
## 24     8791.22110    -1646.358451
## 25     7657.01003    -2819.427732
## 26     9156.81514     9806.356775
## 27    11537.33906      687.011793
## 28    13569.78556    -2593.539810
## 29    11169.07285     -746.156197
## 30     2729.00693     -598.331034
## 31    15718.01687    -2768.861466
## 32     7398.50221    -1086.550214
## 33    14014.47899    -2502.073993
## 34    32317.40205     6391.773950
## 35     8901.96546     1205.255142
## 36      657.87692     2550.910082
## 37    10078.95262    -2660.430619
## 38    31171.20545    -9975.387454
## 39     6860.70244    -2801.989985
## 40    28695.16823     5558.885124
```

```
## 41     2806.84969     1054.359959
## 42    15305.98220    12981.915457
## 43    11879.89589    -3208.704642
## 44    13827.49233     -475.392533
## 45     7699.53491    -2274.511559
## 46    38796.67433     8258.857766
## 47    10607.25081     2349.867192
## 48    30038.56698     6798.900015
## 49     4534.55751     1321.344991
## 50    10668.53136    -3227.030364
## 51     5467.06539      650.429111
## 52     5663.99561    -3166.957306
## 53     3666.46219    -1699.439495
## 54    30812.31531    -9713.761256
## 55    16035.49671    -2115.673811
## 56     9383.82751     -492.688005
## 57    13762.40744    -2914.273137
## 58    12830.06449     -724.744489
## 59    15403.92592    -5603.037717
## 60     9921.49399    -1306.193990
## 61      745.75465     2434.755445
## 62    12536.35420    -4411.945804
## 63    29475.56507    -8729.575971
## 64    30448.17608    -9366.016084
## 65    33124.30866   -10711.660163
## 66     -322.14392     1949.426365
## 67    12902.35530    -4567.897755
## 68     5537.48518    -3620.166780
## 69     9338.06387    -3501.543475
## 70     7386.32664    -2984.093641
## 71    10396.92882    -4107.173920
## 72    10815.43679     1664.272159
## 73    11069.63761    -2463.420210
## 74    37839.42954     7862.592814
## 75     2680.39966     -960.963356
## 76    12972.95947    -2371.327223
## 77    30852.36900    -8373.769001
## 78     9347.33092     1444.629084
## 79    28457.29121    -9416.415212
## 80     9810.66645     1259.868553
## 81    22049.12840    -1881.792369
## 82     7410.31223    -4209.067083
## 83    10706.92963     1322.357070
## 84    13983.94146    -3898.095460
## 85    10471.57177    -3123.429767
## 86     3669.55800    18314.912614
## 87     9842.78926    -3436.378565
## 88      -20.03030     1648.501202
## 89     6210.27461    -1949.530610
## 90    13757.80634    -4709.779039
## 91    25029.64718    -8609.152630
## 92    18017.50752    -4035.657169
## 93     3971.58648    -2709.727481
## 94    39360.57109     7909.282915
```

```
## 95    26989.30512     7790.309879
## 96    26322.98362     7577.669383
## 97     9716.44379    -1112.620385
## 98     5406.84222     1279.589082
## 99    14664.90060    11802.196767
## 100    6494.62479    -3051.560790
## 101   12001.93658    -6292.772175
## 102   39774.28383     8122.507522
## 103    9558.45779    -1247.618638
## 104    8091.39845    -1962.601005
## 105   26928.71907    -9566.952972
## 106   17138.60630    10802.681284
## 107   29620.14443   -10104.602834
## 108   19223.75599    -6631.221494
## 109   15471.34150    -4869.929496
## 110    4471.64083       99.772217
## 111   24480.52541    -8182.679406
## 112     930.09807    24151.669768
## 113    9778.71842    -3030.127219
## 114   27316.29570    -8208.516101
## 115   10357.18336    -2240.503359
## 116   26146.03021     7604.261594
## 117    9716.74938    -4319.132683
## 118   10108.70963     1345.311872
## 119   10166.71730    -5404.388303
## 120   12171.66665    -2393.319453
## 121    6405.96036    -1483.044459
## 122    8951.77680     -649.241152
## 123     886.31054      745.725711
## 124    4432.42877     -894.725768
## 125   10317.51984      765.057356
## 126   29493.07251     5998.567489
## 127    7555.31251    -2180.274506
## 128    3327.82253    -1801.510526
## 129    9555.91837     1386.213677
## 130    4471.11474    -1615.677186
## 131   14453.06302      566.697031
## 132   12761.98971    -1103.610556
## 133   12862.81019      266.793261
## 134    6376.95058     -961.289380
## 135   11219.23012    -4071.757321
## 136    3978.15631      709.640688
## 137   29316.80695     6807.766753
## 138    4997.08059    18244.393939
## 139    4777.61506    -1873.527057
## 140    2298.24797     -633.248375
## 141    4946.17052    -3804.725419
## 142   14043.30139    11290.031448
## 143   10234.03722     1111.481783
## 144   13383.58892    -1151.975318
## 145   12627.70832    -2027.160020
## 146   13041.61526    -3407.077263
## 147   35193.18726     8385.752142
## 148    4367.95644    -1646.635639
```

```
## 149    6918.76606    -2878.207806
## 150   38018.67579     9286.629209
## 151    7011.20164     -887.632837
## 152   12512.00363     -178.175626
## 153    6441.03840    20934.866382
## 154    7806.00713      179.807866
## 155   32223.41866     5384.109039
## 156    1272.16686     1313.683791
## 157    6306.56515    -5159.768554
## 158    7579.66299    -3721.903739
## 159    4069.19606    -2677.667356
## 160   11810.06522    -5449.071615
## 161   13338.77978    -1242.128580
## 162   30267.26691     5921.834790
## 163    6186.65648    -2198.730479
## 164    4821.34275     -581.450095
## 165   11168.08908     1054.809215
## 166   13224.55476      233.406043
## 167   30026.59907    -8252.276922
## 168    1388.19978     2344.425325
## 169   27785.32582    -9526.109822
## 170    9463.10293      400.368872
## 171   14373.62565    -5640.396401
## 172    6198.78552    -1046.651519
## 173    1041.51902      795.717983
## 174   10384.62591    -3026.450263
## 175    7164.61226    -2136.465661
## 176    6292.59147    -3393.102122
## 177   32568.66071     8363.768792
## 178   38896.58621     9927.863786
## 179   16289.08687    -2818.226866
## 180    4492.06127      588.034731
## 181    6013.81137    -3515.396969
## 182    3163.34966     -119.136358
## 183   12709.77157     -385.835573
## 184    8864.36714    -2088.406139
## 185   12359.73769     -805.514094
## 186   25596.11499     6952.225508
## 187   12607.67312     -478.058970
## 188   10114.43050    20170.212439
## 189   15369.46903    -1325.992330
## 190   11371.89671    -4175.029713
## 191   11033.90202    -1751.421424
## 192    6569.63028    -2495.176584
## 193    7212.89790     -166.175698
## 194    5121.28814     -243.307088
## 195   12949.08364    -2541.997791
## 196   13838.24616    -2674.678157
## 197    1529.34753      667.125665
## 198   11601.82142    -1262.889821
## 199   33152.51931    -8758.896909
## 200   40073.26761     8600.291189
## 201    4545.24208    -2898.812381
## 202    3811.20156     7461.129825
```

```
## 203   14159.62066     -314.823460
## 204   11987.22726    -3448.938815
## 205    6548.68641     2645.152089
## 206   16299.35239    20280.929770
## 207    9646.08782     2758.791276
## 208   25998.91317     8440.942730
## 209    1813.46281     1160.663192
## 210    4608.82497    -2470.754274
## 211   27732.91367    -8210.945468
## 212   31718.57102     6992.428983
## 213    7679.90986     -258.715312
## 214   24577.31287    -7492.045272
## 215    6004.79757      267.679628
## 216   10977.14666     1067.195338
## 217    4524.43258    -2273.597376
## 218    7829.69239    -5140.196988
## 219   31420.88602    -9760.955916
## 220    7105.21876     5784.838888
## 221   14782.61379      -89.944437
## 222   16585.32462    -2949.686724
## 223   28528.97193     6277.495769
## 224   16137.68948    -2163.233929
## 225   11271.52744     2179.594565
## 226    5022.94304    -1075.529936
## 227   31512.56788     6230.007818
## 228   10885.97253    -3624.231526
## 229    7853.47163    -1397.608982
## 230   16990.83628     -921.751530
## 231   13190.56474    -3118.509694
## 232    4266.71751    -2246.165210
## 233    6682.62395    -4201.644854
## 234   15904.15648    -3556.984480
## 235   26172.23784    -8675.931845
## 236     -63.68199     1801.057992
## 237    3161.05764     -441.777887
## 238   25047.07534    -9528.895095
## 239    6907.14642     -410.260418
## 240    6117.53381     3526.718686
## 241   16416.63516     7810.702084
## 242   10114.23263      845.097375
## 243    8057.95489     -732.906690
## 244   27363.27972     6369.406985
## 245   11105.40348      414.696365
## 246   12428.84733    -3604.861579
## 247   34174.51194    -7136.597842
## 248   24996.97711     7737.209193
## 249   11074.37553    -1204.565334
## 250   11614.87720      126.848796
## 251    2971.78006      381.503943
## 252   13546.68403    -2721.430334
## 253   10073.94029     2020.537705
## 254    3663.51272    -2527.113321
## 255   13300.88765      700.399046
## 256    5405.13248    -1760.043075
```

```
## 257    5387.83712    -3753.263716
## 258    9996.08724    -1715.464538
## 259    7569.10393    18667.476036
## 260   34291.66918     6428.881866
## 261    6350.30470     -151.552904
## 262   18571.18777    -2486.060274
## 263    9126.12163    -5136.280625
## 264   31109.27336    -9885.597563
## 265    9097.51625     8781.384428
## 266    7327.78583    -1358.062830
## 267    3746.87584     1184.771160
## 268   11465.82893    -2630.563976
## 269    4927.66088    -1542.261730
## 270    4688.20559     9444.832160
## 271    2523.70322      -59.084415
## 272   11840.65683     6965.488639
## 273    1390.46614      579.147860
## 274    9445.86486     8483.438506
## 275   25276.00844    -7052.557239
## 276   10606.84695    -4132.833949
## 277   34494.87447    -8385.545422
## 278    2692.99668     -555.343076
## 279   35753.46531     5343.696441
## 280   28352.35541    -8418.897406
## 281   10582.62776     2429.580893
## 282   11805.47434    -1573.974438
## 283   18782.17450    -4371.242402
## 284   16353.90083    -3210.564185
## 285    2158.42843     1101.770568
## 286   32147.63221     7688.886794
## 287   22442.57588    -7730.832079
## 288   14154.21743    -1596.612128
## 289    8270.76036    -1988.525363
## 290   11502.34667    -3479.211224
## 291   10114.06716     3090.218488
## 292   37438.05512    -6495.863325
## 293   10668.49973    -1221.249376
## 294   12753.28031    -3789.219757
## 295   33239.30945     6632.394848
## 296    8407.50271    -5138.656062
## 297    8273.38408    -1414.904484
## 298   13846.00983    -4867.824728
## 299   30437.71782    20756.841321
## 300    9629.41690    -2117.149903
```

```
#-----------------------------------------------------------------------
# Both fitted values and residual values match with matrix model
lm(charges ~ age + children + bmi+ region + sex + smoker, data= insurance_new) -> AB
summary(AB)
```

```
##
## Call:
## lm(formula = charges ~ age + children + bmi + region + sex +
##     smoker, data = insurance_new)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -10712  -3120  -1095   1496  24152
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -12033.25    2155.91  -5.582 5.46e-08 ***
## age                261.14      24.05  10.858  < 2e-16 ***
## children           532.76     279.27   1.908   0.0574 .
## bmi                353.35      62.03   5.696 2.99e-08 ***
## regionnorthwest  -1545.53     992.47  -1.557   0.1205
## regionsoutheast  -1505.35    1036.22  -1.453   0.1474
## regionsouthwest  -1719.90     990.21  -1.737   0.0835 .
## sexmale            607.78     694.59   0.875   0.3823
## smokeryes        22876.38     865.37  26.435  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5915 on 291 degrees of freedom
## Multiple R-squared:  0.7379, Adjusted R-squared:  0.7307
## F-statistic: 102.4 on 8 and 291 DF,  p-value: < 2.2e-16
```

```
residuals(AB)
```

```
##            1            2            3            4            5
##   2519.297588   8174.301498  11383.487674  -2582.201071  -1913.506557
##            6            7            8            9           10
##   1074.351480   2097.967229  15414.919004      6.065788   3244.090940
##           11           12           13           14           15
##  -1224.929878  11983.339430   2180.251898   3265.176147  -3189.320658
##           16           17           18           19           20
##  -1935.611645    951.540493  -6807.332617   4498.696808   -922.722854
##           21           22           23           24           25
##    305.939817  -8424.114330   -944.151815  -1646.358451  -2819.427732
##           26           27           28           29           30
##   9806.356775    687.011793  -2593.539810   -746.156197   -598.331034
##           31           32           33           34           35
##  -2768.861466  -1086.550214  -2502.073993   6391.773950   1205.255142
##           36           37           38           39           40
##   2550.910082  -2660.430619  -9975.387454  -2801.989985   5558.885124
##           41           42           43           44           45
##   1054.359959  12981.915457  -3208.704642   -475.392533  -2274.511559
##           46           47           48           49           50
##   8258.857766   2349.867192   6798.900015   1321.344991  -3227.030364
##           51           52           53           54           55
##    650.429111  -3166.957306  -1699.439495  -9713.761256  -2115.673811
##           56           57           58           59           60
##   -492.688005  -2914.273137   -724.744489  -5603.037717  -1306.193990
##           61           62           63           64           65
##   2434.755445  -4411.945804  -8729.575971  -9366.016084 -10711.660163
##           66           67           68           69           70
##   1949.426365  -4567.897755  -3620.166780  -3501.543475  -2984.093641
##           71           72           73           74           75
```

```
##  -4107.173920   1664.272159  -2463.420210   7862.592814   -960.963356
##           76            77            78            79            80
##  -2371.327223  -8373.769001   1444.629084  -9416.415212   1259.868553
##           81            82            83            84            85
##  -1881.792369  -4209.067083   1322.357070  -3898.095460  -3123.429767
##           86            87            88            89            90
##  18314.912614  -3436.378565   1648.501202  -1949.530610  -4709.779039
##           91            92            93            94            95
##  -8609.152630  -4035.657169  -2709.727481   7909.282915   7790.309879
##           96            97            98            99           100
##   7577.669383  -1112.620385   1279.589082  11802.196767  -3051.560790
##          101           102           103           104           105
##  -6292.772175   8122.507522  -1247.618638  -1962.601005  -9566.952972
##          106           107           108           109           110
##  10802.681284 -10104.602834  -6631.221494  -4869.929496     99.772217
##          111           112           113           114           115
##  -8182.679406  24151.669768  -3030.127219  -8208.516101  -2240.503359
##          116           117           118           119           120
##   7604.261594  -4319.132683   1345.311872  -5404.388303  -2393.319453
##          121           122           123           124           125
##  -1483.044459   -649.241152    745.725711   -894.725768    765.057356
##          126           127           128           129           130
##   5998.567489  -2180.274506  -1801.510526   1386.213677  -1615.677186
##          131           132           133           134           135
##    566.697031  -1103.610556    266.793261   -961.289380  -4071.757321
##          136           137           138           139           140
##    709.640688   6807.766753  18244.393939  -1873.527057   -633.248375
##          141           142           143           144           145
##  -3804.725419  11290.031448   1111.481783  -1151.975318  -2027.160020
##          146           147           148           149           150
##  -3407.077263   8385.752142  -1646.635639  -2878.207806   9286.629209
##          151           152           153           154           155
##   -887.632837   -178.175626  20934.866382    179.807866   5384.109039
##          156           157           158           159           160
##   1313.683791  -5159.768554  -3721.903739  -2677.667356  -5449.071615
##          161           162           163           164           165
##  -1242.128580   5921.834790  -2198.730479   -581.450095   1054.809215
##          166           167           168           169           170
##    233.406043  -8252.276922   2344.425325  -9526.109822    400.368872
##          171           172           173           174           175
##  -5640.396401  -1046.651519    795.717983  -3026.450263  -2136.465661
##          176           177           178           179           180
##  -3393.102122   8363.768792   9927.863786  -2818.226866    588.034731
##          181           182           183           184           185
##  -3515.396969   -119.136358   -385.835573  -2088.406139   -805.514094
##          186           187           188           189           190
##   6952.225508   -478.058970  20170.212439  -1325.992330  -4175.029713
##          191           192           193           194           195
##  -1751.421424  -2495.176584   -166.175698   -243.307088  -2541.997791
##          196           197           198           199           200
##  -2674.678157    667.125665  -1262.889821  -8758.896909   8600.291189
##          201           202           203           204           205
##  -2898.812381   7461.129825   -314.823460  -3448.938815   2645.152089
##          206           207           208           209           210
```

```
##  20280.929770   2758.791276   8440.942730   1160.663192  -2470.754274
##           211           212           213           214           215
##  -8210.945468   6992.428983   -258.715312  -7492.045272    267.679628
##           216           217           218           219           220
##   1067.195338  -2273.597376  -5140.196988  -9760.955916   5784.838888
##           221           222           223           224           225
##    -89.944437  -2949.686724   6277.495769  -2163.233929   2179.594565
##           226           227           228           229           230
##  -1075.529936   6230.007818  -3624.231526  -1397.608982   -921.751530
##           231           232           233           234           235
##  -3118.509694  -2246.165210  -4201.644854  -3556.984480  -8675.931845
##           236           237           238           239           240
##   1801.057992   -441.777887  -9528.895095   -410.260418   3526.718686
##           241           242           243           244           245
##   7810.702084    845.097375   -732.906690   6369.406985    414.696365
##           246           247           248           249           250
##  -3604.861579  -7136.597842   7737.209193  -1204.565334    126.848796
##           251           252           253           254           255
##    381.503943  -2721.430334   2020.537705  -2527.113321    700.399046
##           256           257           258           259           260
##  -1760.043075  -3753.263716  -1715.464538  18667.476036   6428.881866
##           261           262           263           264           265
##   -151.552904  -2486.060274  -5136.280625  -9885.597563   8781.384428
##           266           267           268           269           270
##  -1358.062830   1184.771160  -2630.563976  -1542.261730   9444.832160
##           271           272           273           274           275
##    -59.084415   6965.488639    579.147860   8483.438506  -7052.557239
##           276           277           278           279           280
##  -4132.833949  -8385.545422   -555.343076   5343.696441  -8418.897406
##           281           282           283           284           285
##   2429.580893  -1573.974438  -4371.242402  -3210.564185   1101.770568
##           286           287           288           289           290
##   7688.886794  -7730.832079  -1596.612128  -1988.525363  -3479.211224
##           291           292           293           294           295
##   3090.218488  -6495.863325  -1221.249376  -3789.219757   6632.394848
##           296           297           298           299           300
##  -5138.656062  -1414.904484  -4867.824728  20756.841321  -2117.149903
```

```
fitted(AB)
```

```
##           1           2           3           4           5           6
## 11735.31061 28133.49680  9848.69459 10822.79067 11638.03656  4383.69497
##           7           8           9          10          11          12
##  -265.87323 13054.00001  1738.39921 -1512.41394 10850.84988  6174.53657
##          13          14          15          16          17          18
##  9653.53040  8767.14985  5016.16366 15758.41465  3577.93651 12246.08172
##          19          20          21          22          23          24
##  9979.63334  2800.65225  1398.62828 32605.04783  7994.79382  8791.22110
##          25          26          27          28          29          30
##  7657.01003  9156.81514 11537.33906 13569.78556 11169.07285  2729.00693
##          31          32          33          34          35          36
## 15718.01687  7398.50221 14014.47899 32317.40205  8901.96546   657.87692
##          37          38          39          40          41          42
## 10078.95262 31171.20545  6860.70244 28695.16823  2806.84969 15305.98220
```

```
##          43          44          45          46          47          48
## 11879.89589 13827.49233  7699.53491 38796.67433 10607.25081 30038.56698
##          49          50          51          52          53          54
##  4534.55751 10668.53136  5467.06539  5663.99561  3666.46219 30812.31531
##          55          56          57          58          59          60
## 16035.49671  9383.82751 13762.40744 12830.06449 15403.92592  9921.49399
##          61          62          63          64          65          66
##   745.75465 12536.35420 29475.56507 30448.17608 33124.30866  -322.14392
##          67          68          69          70          71          72
## 12902.35530  5537.48518  9338.06387  7386.32664 10396.92882 10815.43679
##          73          74          75          76          77          78
## 11069.63761 37839.42954  2680.39966 12972.95947 30852.36900  9347.33092
##          79          80          81          82          83          84
## 28457.29121  9810.66645 22049.12840  7410.31223 10706.92963 13983.94146
##          85          86          87          88          89          90
## 10471.57177  3669.55800  9842.78926   -20.03030  6210.27461 13757.80634
##          91          92          93          94          95          96
## 25029.64718 18017.50752  3971.58648 39360.57109 26989.30512 26322.98362
##          97          98          99         100         101         102
##  9716.44379  5406.84222 14664.90060  6494.62479 12001.93658 39774.28383
##         103         104         105         106         107         108
##  9558.45779  8091.39845 26928.71907 17138.60630 29620.14443 19223.75599
##         109         110         111         112         113         114
## 15471.34150  4471.64083 24480.52541   930.09807  9778.71842 27316.29570
##         115         116         117         118         119         120
## 10357.18336 26146.03021  9716.74938 10108.70963 10166.71730 12171.66665
##         121         122         123         124         125         126
##  6405.96036  8951.77680   886.31054  4432.42877 10317.51984 29493.07251
##         127         128         129         130         131         132
##  7555.31251  3327.82253  9555.91837  4471.11474 14453.06302 12761.98971
##         133         134         135         136         137         138
## 12862.81019  6376.95058 11219.23012  3978.15631 29316.80695  4997.08059
##         139         140         141         142         143         144
##  4777.61506  2298.24797  4946.17052 14043.30139 10234.03722 13383.58892
##         145         146         147         148         149         150
## 12627.70832 13041.61526 35193.18726  4367.95644  6918.76606 38018.67579
##         151         152         153         154         155         156
##  7011.20164 12512.00363  6441.03840  7806.00713 32223.41866  1272.16686
##         157         158         159         160         161         162
##  6306.56515  7579.66299  4069.19606 11810.06522 13338.77978 30267.26691
##         163         164         165         166         167         168
##  6186.65648  4821.34275 11168.08908 13224.55476 30026.59907  1388.19978
##         169         170         171         172         173         174
## 27785.32582  9463.10293 14373.62565  6198.78552  1041.51902 10384.62591
##         175         176         177         178         179         180
##  7164.61226  6292.59147 32568.66071 38896.58621 16289.08687  4492.06127
##         181         182         183         184         185         186
##  6013.81137  3163.34966 12709.77157  8864.36714 12359.73769 25596.11499
##         187         188         189         190         191         192
## 12607.67312 10114.43050 15369.46903 11371.89671 11033.90202  6569.63028
##         193         194         195         196         197         198
##  7212.89790  5121.28814 12949.08364 13838.24616  1529.34753 11601.82142
##         199         200         201         202         203         204
## 33152.51931 40073.26761  4545.24208  3811.20156 14159.62066 11987.22726
```

```
##         205         206         207         208         209         210
##  6548.68641 16299.35239  9646.08782 25998.91317  1813.46281  4608.82497
##         211         212         213         214         215         216
## 27732.91367 31718.57102  7679.90986 24577.31287  6004.79757 10977.14666
##         217         218         219         220         221         222
##  4524.43258  7829.69239 31420.88602  7105.21876 14782.61379 16585.32462
##         223         224         225         226         227         228
## 28528.97193 16137.68948 11271.52744  5022.94304 31512.56788 10885.97253
##         229         230         231         232         233         234
##  7853.47163 16990.83628 13190.56474  4266.71751  6682.62395 15904.15648
##         235         236         237         238         239         240
## 26172.23784   -63.68199  3161.05764 25047.07534  6907.14642  6117.53381
##         241         242         243         244         245         246
## 16416.63516 10114.23263  8057.95489 27363.27972 11105.40348 12428.84733
##         247         248         249         250         251         252
## 34174.51194 24996.97711 11074.37553 11614.87720  2971.78006 13546.68403
##         253         254         255         256         257         258
## 10073.94029  3663.51272 13300.88765  5405.13248  5387.83712  9996.08724
##         259         260         261         262         263         264
##  7569.10393 34291.66918  6350.30470 18571.18777  9126.12163 31109.27336
##         265         266         267         268         269         270
##  9097.51625  7327.78583  3746.87584 11465.82893  4927.66088  4688.20559
##         271         272         273         274         275         276
##  2523.70322 11840.65683  1390.46614  9445.86486 25276.00844 10606.84695
##         277         278         279         280         281         282
## 34494.87447  2692.99668 35753.46531 28352.35541 10582.62776 11805.47434
##         283         284         285         286         287         288
## 18782.17450 16353.90083  2158.42843 32147.63221 22442.57588 14154.21743
##         289         290         291         292         293         294
##  8270.76036 11502.34667 10114.06716 37438.05512 10668.49973 12753.28031
##         295         296         297         298         299         300
## 33239.30945  8407.50271  8273.38408 13846.00983 30437.71782  9629.41690
```

5 Analyze and evaluate the full model

```
lm(charges ~ age + children + bmi+ region + sex + smoker, data= insurance_new) -> AB
summary(AB)
```

```
##
## Call:
## lm(formula = charges ~ age + children + bmi + region + sex +
##     smoker, data = insurance_new)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -10712  -3120  -1095   1496  24152
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -12033.25    2155.91  -5.582 5.46e-08 ***
## age                261.14      24.05  10.858  < 2e-16 ***
```

```
## children           532.76     279.27   1.908   0.0574 .
## bmi                353.35      62.03   5.696 2.99e-08 ***
## regionnorthwest  -1545.53     992.47  -1.557   0.1205
## regionsoutheast  -1505.35    1036.22  -1.453   0.1474
## regionsouthwest  -1719.90     990.21  -1.737   0.0835 .
## sexmale            607.78     694.59   0.875   0.3823
## smokeryes        22876.38     865.37  26.435  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5915 on 291 degrees of freedom
## Multiple R-squared:  0.7379, Adjusted R-squared:  0.7307
## F-statistic: 102.4 on 8 and 291 DF,  p-value: < 2.2e-16
```

5.1 Coefficients:

charges = -12033.25 + (age)*261.14 + (children)*532.76 + (bmi)*353.35 + (region_northwest)*(-1545.53) + (region_southeast)*(-1505.35) + (regionsouthwest)*(-1719.90) + (sexmale)*(607.78) + (smokeryes)*(22876.38)

The **Estimate column** lists the estimated coefficients of the predictor variables included in the model. For instance, the "age" coefficient has an estimated value of 261.14, which means that for every one-unit increase in age, the outcome variable (presumably a medical cost) is estimated to increase by $261.14, holding all other variables constant.

5.2 Standard Errors:

The **Std Error** column lists the standard errors of the estimated coefficients. These are measures of the uncertainty or variability in the estimated coefficients. Smaller standard errors indicate more precise estimates.

5.3 T-values:

The **t-value** column lists the t-statistics for each coefficient. These values represent the estimated coefficients divided by their standard errors. T-values are used to test the null hypothesis that the true coefficient is zero. Larger t-values indicate a stronger evidence against the null hypothesis.

5.4 P-values

The pr column lists the p-values associated with the t-values. P-values represent the probability of observing the t-value or a more extreme value if the true coefficient is zero. Smaller p-values indicate stronger evidence against the null hypothesis.

The significance codes provided in the table help to quickly identify significant coefficients; for instance, **age,bmi,smokers** represents p-value less than two decimal places and **children** represents p-value less than 0.05.One can say that **region** represents bigger p-value than significant level.

5.5 Residual Standard Error:

The **Residual standard error** provides an estimate of the variability of the errors or unexplained variance in the model. It measures the average distance that the observed values fall from the predicted values.

5.6 R-squared:

The "Multiple R-squared" and "Adjusted R-squared" measures how well the model fits the data. R-squared ranges from 0 to 1 and represents the proportion of variance in the outcome variable that is explained by the predictor variables. Adjusted R-squared is a corrected version of R-squared that takes into account the number of predictor variables in the model. In full model one can see 0.7379.

5.7 summary

This data summary provides information on the estimated coefficients, their standard errors, t-values, and corresponding p-values, as well as model diagnostics such as residual standard error, multiple R-squared, adjusted R-squared, and F-statistic. These measures can be used to interpret the strength and significance of the relationship between the predictor variables and the outcome variable, as well as the overall goodness-of-fit of the model. Still working to find the best parameters to be included in the model.

6 confidence intervals for all variables used in full model

```
confint(AB,level = 0.95)
```

```
##                       2.5 %     97.5 %
## (Intercept)     -16276.39288 -7790.1053
## age               213.80684   308.4817
## children          -16.88546  1082.3959
## bmi               231.26578   475.4327
## regionnorthwest  -3498.86525   407.8049
## regionsoutheast  -3544.78722   534.0830
## regionsouthwest  -3668.78747   228.9853
## sexmale           -759.27612  1974.8375
## smokeryes        21173.19777 24579.5601
```

7 Now produce a reduced model (removing variables of your choice with justification). Use R summary coding for both models and offer justification for choosing one model over the othe

7.1 Evaluating Various regression models.

Let us begin by using a multiple linear regression model that uses all the six variables. From the summary table we see that our R squared and Adjusted r square are around 0.73 and the residual standard error is 5915. The r squared value is high enough to be considered good but let us continue finding better fits.

```
ols_step_all_possible(AB) ->  allmodels
as_tibble(allmodels) -> allmodels_1
tail(allmodels_1,8)
```

```
## # A tibble: 8 x 14
##   mindex     n predict~1 rsquare   adjr predrsq     cp   aic  sbic   sbc    msep
```

```
##    <int> <int> <chr>       <dbl>  <dbl>   <dbl>  <dbl> <dbl> <dbl> <dbl>   <dbl>
## 1     56     4 children~  0.0470 0.0275 0.00150 768.   6457. 5593. 6486. 3.77e10
## 2     57     5 age chil~  0.737  0.731  0.721     3.77 6072. 5217. 6105. 1.04e10
## 3     58     5 age bmi ~  0.735  0.728  0.718     6.64 6075. 5220. 6108. 1.05e10
## 4     59     5 age chil~  0.735  0.730  0.722     6.70 6071. 5220. 6097. 1.05e10
## 5     60     5 age chil~  0.709  0.702  0.690    35.4  6103. 5247. 6136. 1.15e10
## 6     61     5 children~  0.632  0.623  0.609   121.   6173. 5315. 6207. 1.46e10
## 7     62     5 age chil~  0.109  0.0873 0.0600  702.   6439. 5574. 6472. 3.53e10
## 8     63     6 age chil~  0.738  0.731  0.720     5    6073. 5218. 6110. 1.04e10
## # ... with 3 more variables: fpe <dbl>, apc <dbl>, hsp <dbl>, and abbreviated
## #   variable name 1: predictors
```

```
# the best model is represented by including all age, children,bmi, region, smoker full model.
#plot(allmodels)
```