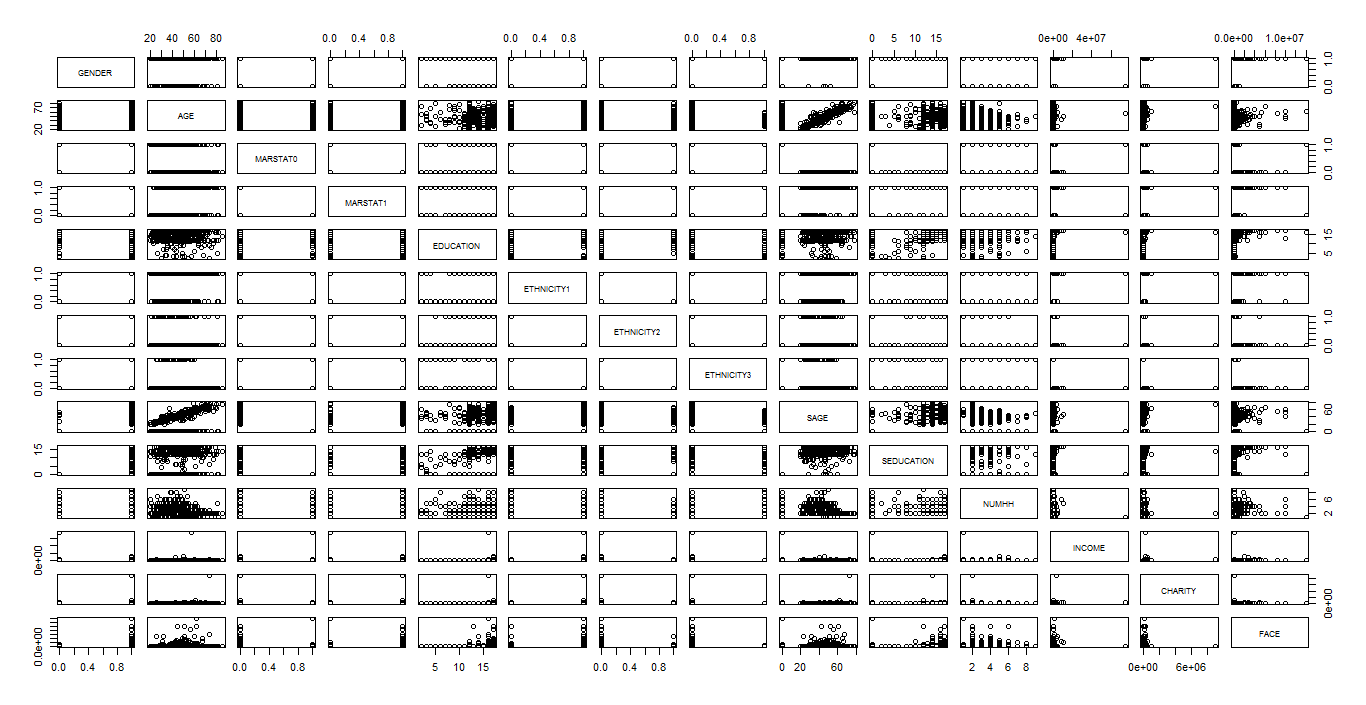
Objective is to build a model to predict the amount that the company will pay in the event of the death of the named insured based on the given variables.

Before proceeding with the analysis, the variables MARSTAT and Ethnicity are categorical variables and need to create some dummy variables for these.

Total number of dummies for k level variable is (k-1)

Imported the data in to the R

data<-read.csv(file.choose())

**1) Scatter plot:**

pairs(data[,1:14])

From the above plot we see that AGE and SAGE are correlated.

**Correlation Matrix:**

cor(data[,1:14])

From the above matrix it is evident that there are some variable are correlated and MARSTAT and SEDUCATION are highly correlated. And there are moderately high correlated variable SAGEVs MARSTAT0.

**Regression model**

m1=lm(FACE~.,data=data)

summary(m1)

Call:

lm(formula = FACE ~ ., data = data)

Residuals:

Min 1Q Median 3Q Max

-1029208 -483191 -225964 39265 13421598

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.835e+06 6.184e+05 -2.968 0.00315 \*\*

GENDER 2.796e+05 2.155e+05 1.297 0.19512

AGE 9.020e+03 6.495e+03 1.389 0.16555

MARSTAT0 5.329e+05 5.397e+05 0.987 0.32391

MARSTAT1 7.916e+04 2.512e+05 0.315 0.75284

EDUCATION 6.515e+04 2.452e+04 2.657 0.00813 \*\*

ETHNICITY1 -1.013e+05 2.624e+05 -0.386 0.69955

ETHNICITY2 -2.198e+05 3.013e+05 -0.729 0.46606

ETHNICITY3 -1.252e+05 3.496e+05 -0.358 0.72047

SAGE -2.946e+03 8.147e+03 -0.362 0.71783

SEDUCATION 4.128e+04 2.539e+04 1.626 0.10464

NUMHH 9.567e+04 4.632e+04 2.065 0.03943 \*

INCOME 7.961e-03 1.674e-02 0.476 0.63461

CHARITY -2.948e-02 1.403e-01 -0.210 0.83372

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1261000 on 486 degrees of freedom

Multiple R-squared: 0.07676, Adjusted R-squared: 0.05206

F-statistic: 3.108 on 13 and 486 DF, p-value: 0.0001824

Ran the regression model with all the variables and results are not at all satisfactory. Only two variables and intercept are significant and rest all are non significant. VIF will be given by

**VIF**

vif(m1)

GENDER AGE MARSTAT0 MARSTAT1 EDUCATION ETHNICITY1 ETHNICITY2 ETHNICITY3

2.098016 2.538102 **18.130836** 4.413349 1.632047 4.266933 3.435961 2.828364

SAGE SEDUCATION NUMHH INCOME CHARITY

**11.220057** 9.029123 1.503528 1.022859 1.028865

From the above VIF it can be inferred that only SAGE and MARSTAT0 variables are having VIF>10. May be at this time we can say that not much information is available to say anything about collinearity.

**Deletion Diagnostics**

influence.measures(m1)

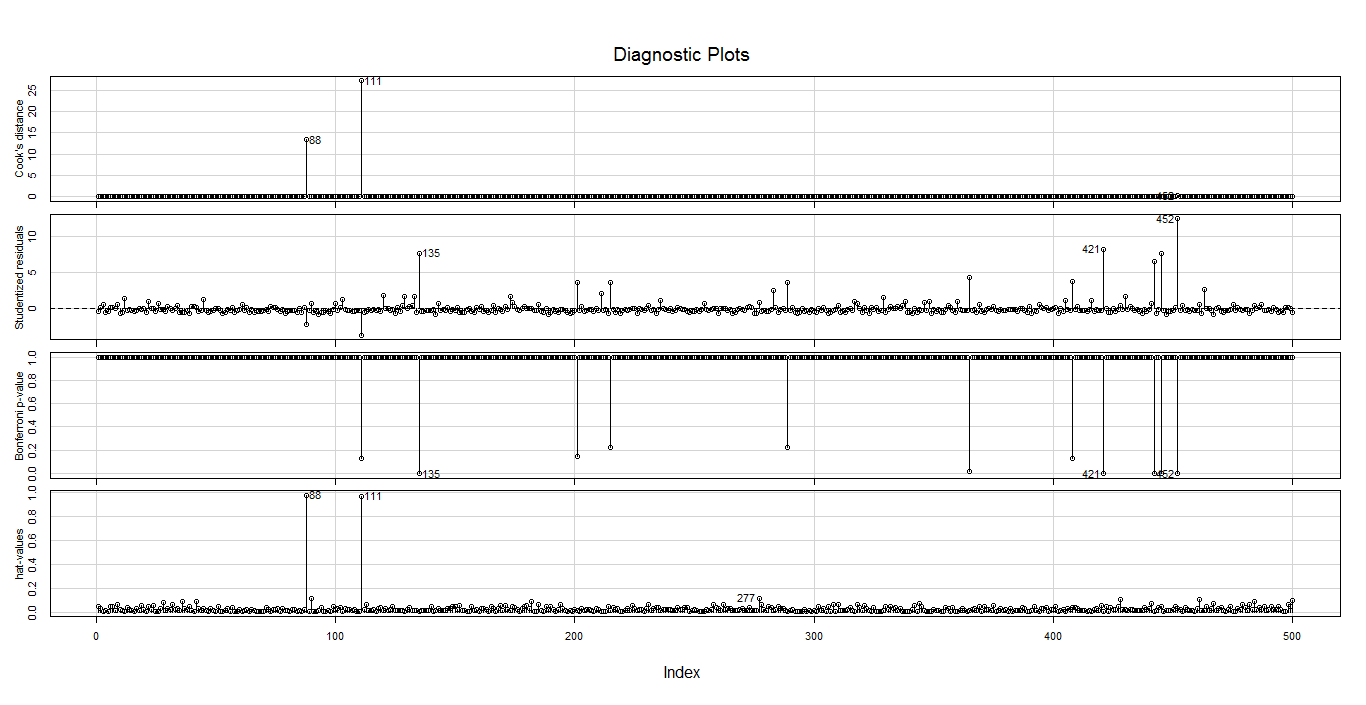
dffit cov.r cook.d hat inf

1.45e-01 1.0908 1.51e-03 0.06471 \*

6.52e-03 1.0935 3.04e-06 0.05878 \*

The above observations are influential observations.

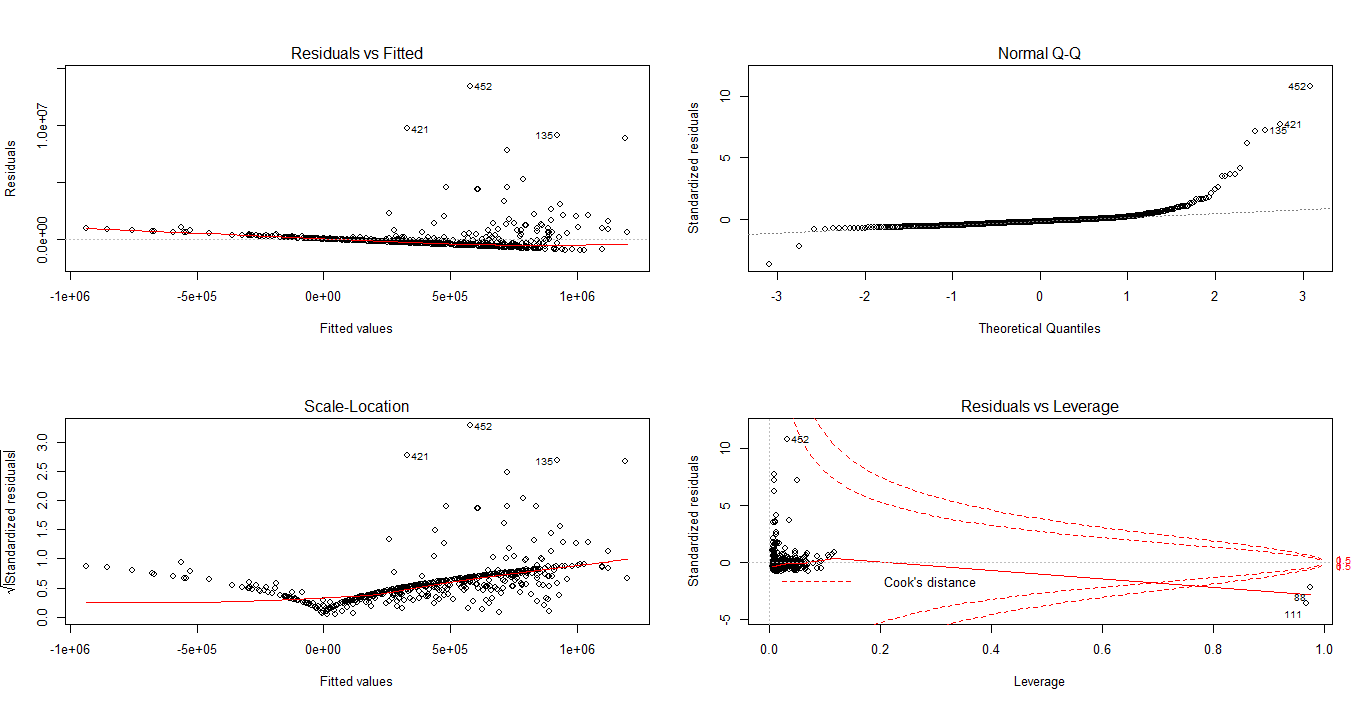
infIndexPlot (m2) //Package(alr3)



Above plot shows that, based on the cook's distance observations 88and 111 are influential observations and from standardized residuals, it is observed that observations 135,452 and 421 are outliers

par(mfrow=c(2,2))

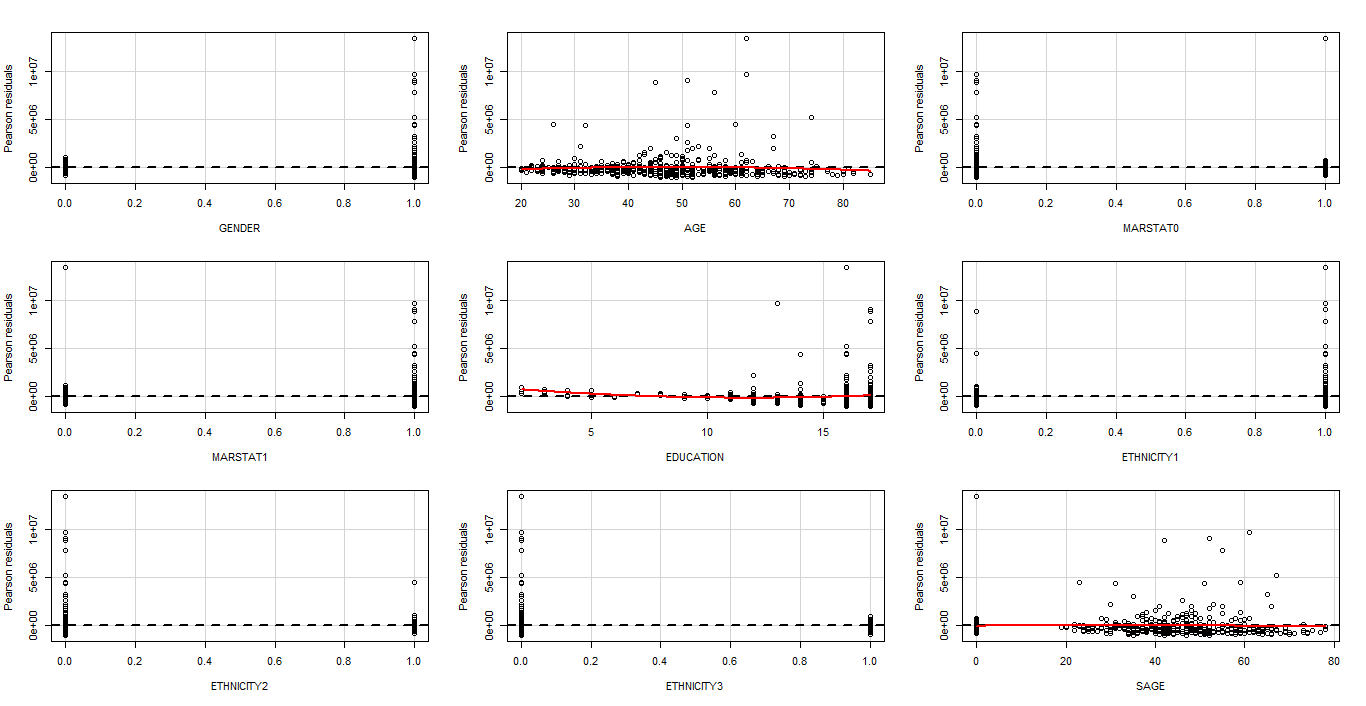
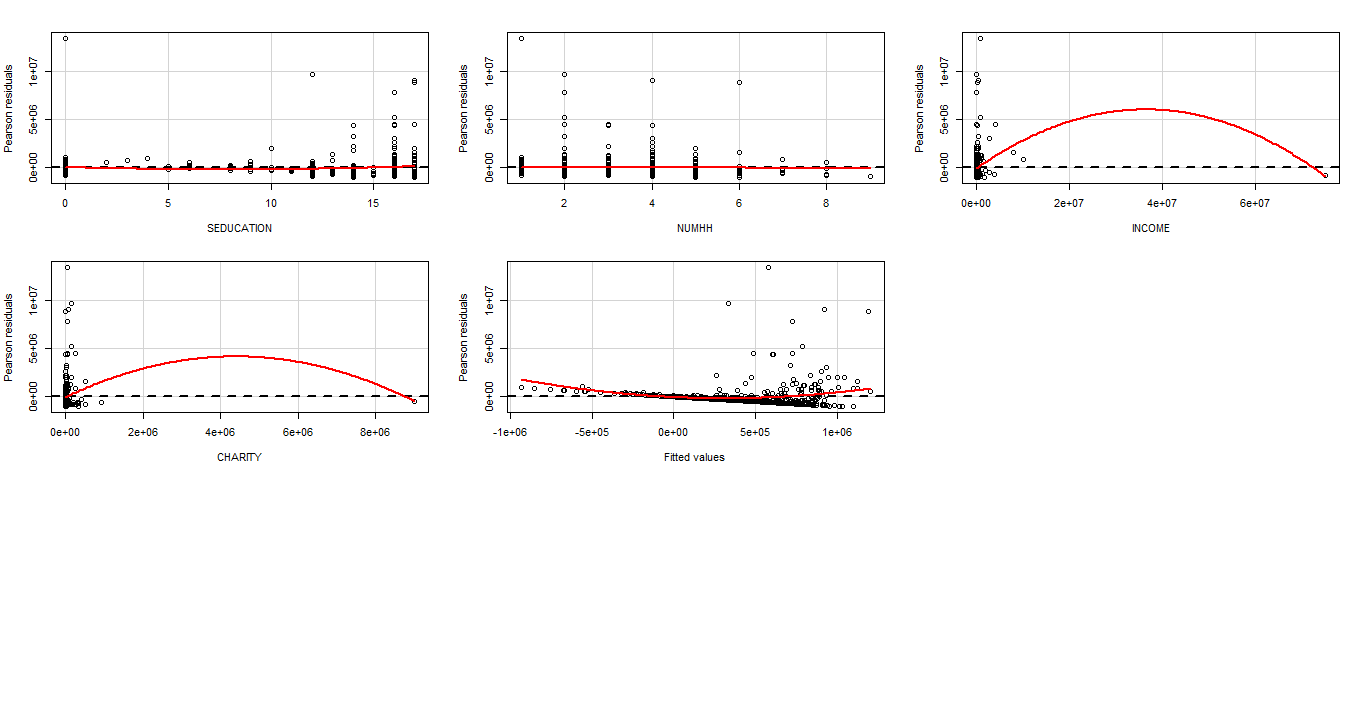
plot(m1)



From the residuals Vs fitted, variance is not constant and data is not normal. From Scale location plot, errors are increasing with fitted value increase.

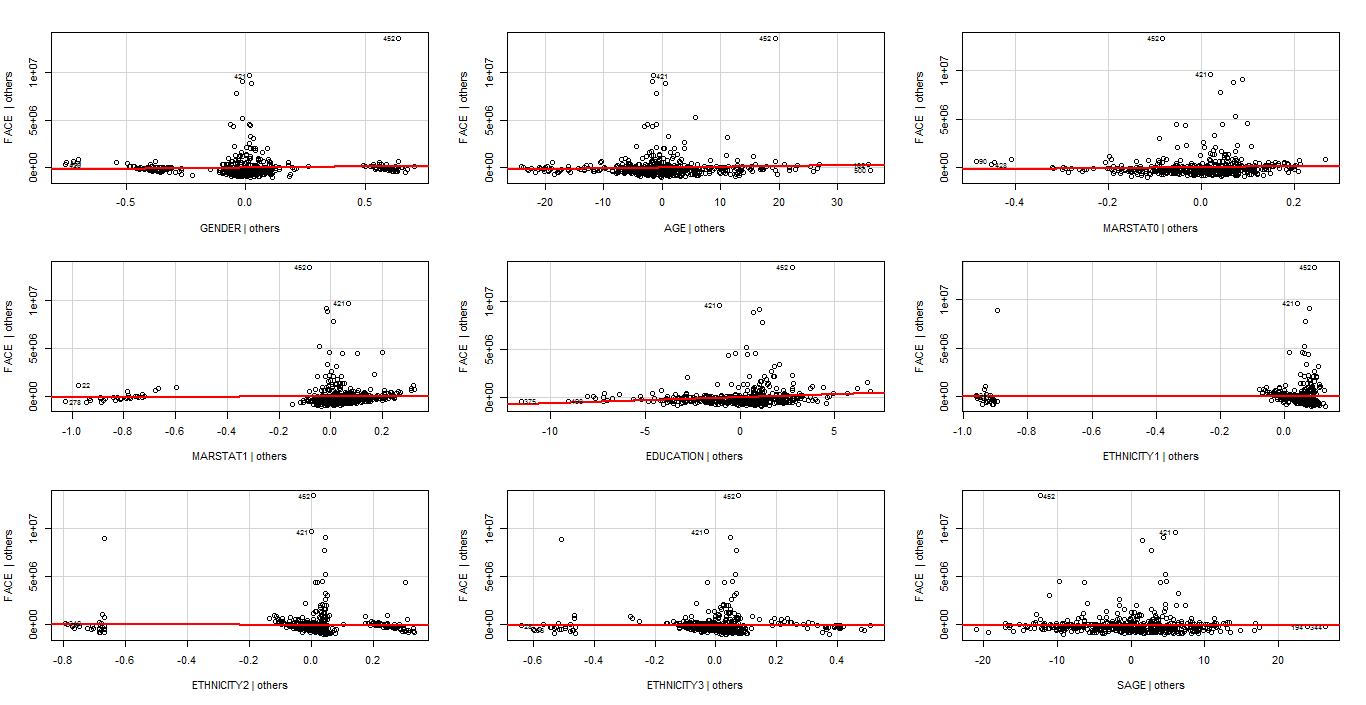
From the residual Vs Leverage, there are some observations which have high errors such as 452 and for 88,111 leverage is high also it crossed the cook's distance line which means those are influential observations.

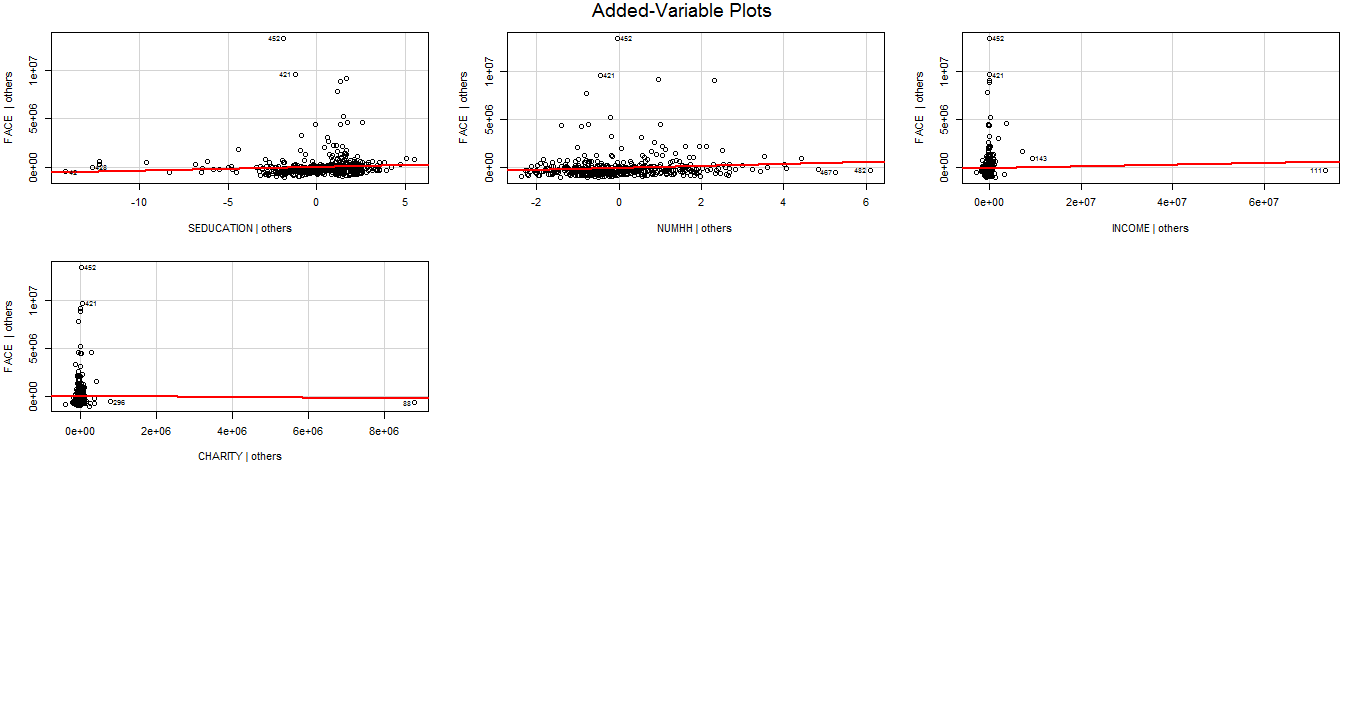
**residualPlots(m1)**



Above plots are regressors Vs residual which will reveal the necessary transformations. Plot **Income** and **Charity** have quadratic distribution and require square transformation.

avPlots(m1, id.n=2, id.cex=0.7)





Added variable plots helps in identifying the real contributor to the response variable and can take decisions such as which variables to be excluded from the model.

From the table we can observe that no variable is in collinear but if we see the intercept it is >0.993.Here we can think of centering the variable to remove the ill effects collinearity.

###################################################################################

**From the residual Vs regressors plots there is need to transform the variables INCOME and CHARITY as they exhibits quadratic nature**

data$sqrincome <- as.numeric(data$INCOME \* data$INCOME)

data$sqrcharity <- as.numeric(data$CHARITY \* data$CHARITY)

Running the regression on the transformed variables.

m3 <- lm(FACE~.,data=data)

> summary(m3)

Call:

lm(formula = FACE ~ ., data = data)

Residuals:

Min 1Q Median 3Q Max

-2021514 -436502 -209383 40942 13165037

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.517e+06 6.176e+05 -2.457 0.01437 \*

GENDER 2.632e+05 2.126e+05 1.238 0.21625

AGE 7.943e+03 6.412e+03 1.239 0.21605

MARSTAT0 4.243e+05 5.334e+05 0.795 0.42672

MARSTAT1 6.391e+04 2.478e+05 0.258 0.79658

EDUCATION 5.497e+04 2.438e+04 2.254 0.02461 \*

ETHNICITY1 -1.059e+05 2.594e+05 -0.408 0.68333

ETHNICITY2 -2.349e+05 2.976e+05 -0.789 0.43044

ETHNICITY3 -1.397e+05 3.456e+05 -0.404 0.68612

SAGE -3.023e+03 8.080e+03 -0.374 0.70844

SEDUCATION 3.276e+04 2.514e+04 1.303 0.19324

NUMHH 8.217e+04 4.584e+04 1.792 0.07369 .

INCOME 3.463e-01 1.117e-01 3.100 0.00205 \*\*

CHARITY 6.902e-01 1.033e+00 0.668 0.50428

Sqrincome -4.692e-09 1.489e-09 -3.152 0.00172 \*\*

Sqrcharity -9.516e-08 1.136e-07 -0.837 0.40278

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1244000 on 484 degrees of freedom

Multiple R-squared: 0.1057, Adjusted R-squared: 0.07795

F-statistic: 3.812 on 15 and 484 DF, p-value: 1.952e-06

Above results are not encouraging w.r.to R square and variable significance. Transformed variable Sqrincome is significant, maybe we can keep that variable for our analysis

GENDER AGE MARSTAT0 MARSTAT1 EDUCATION ETHNICITY1 ETHNICITY2 ETHNICITY3

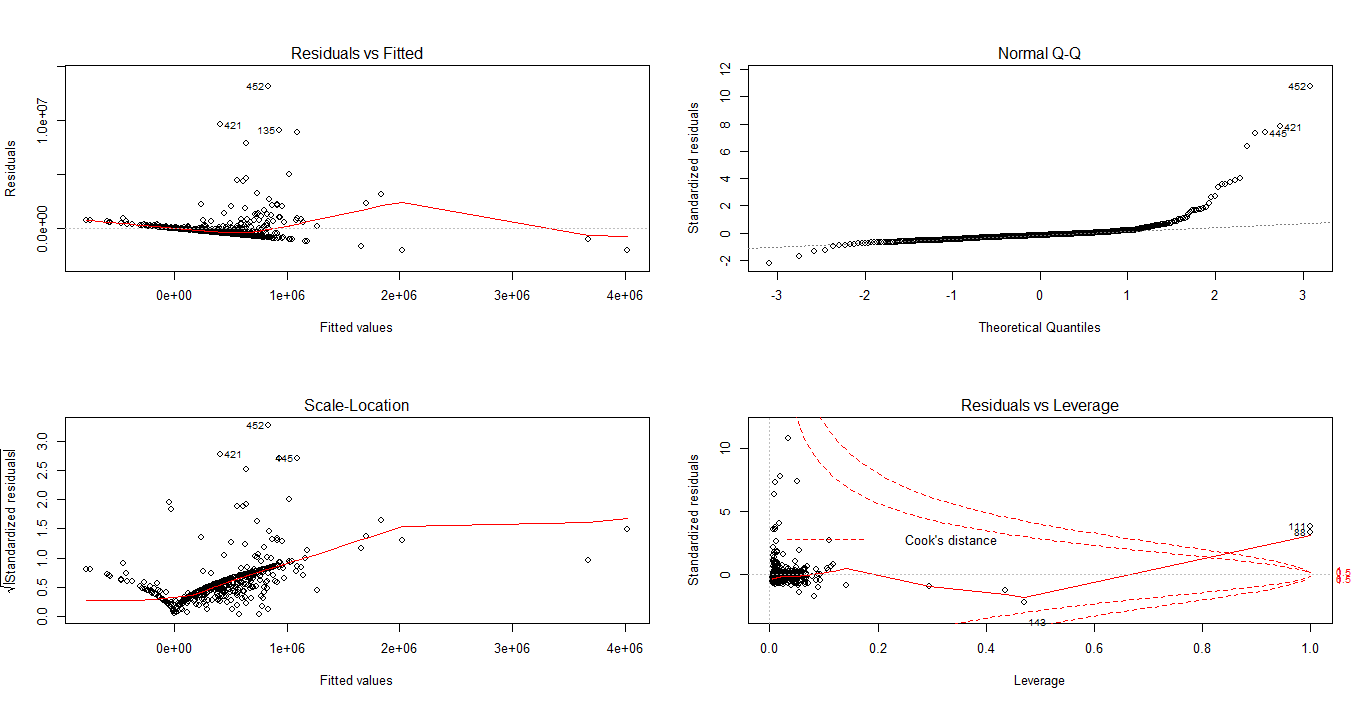
2.099502 2.542886 18.203937 4.414426 1.659889 4.286066 3.446523 2.840224

SAGE SEDUCATION NUMHH INCOME CHARITY Sqrincome Sqrcharity

**11.347148** 9.104380 1.513721 **46.834413** **57.283280** **45.237705** **54.899163**

From the VIF it is clear that there is chance of collinearity exists in the data. To confirm this need to run the variance decomposition proportion table.

plot(m3)



From the above plots and Rsqare values, results are not satisfactory and need to analyse the data further.

From the above plots and influence index plot, there are some outliers which are not influential and can be removed one by one by examining the confidents.

> m4 <- lm(FACE~GENDER+AGE+MARSTAT0+MARSTAT1+EDUCATION+ETHNICITY1+ETHNICITY2+ETHNICITY3+SAGE+SEDUCATION+NUMHH+INCOME+Sqrincome+CHARITY+Sqrcharity, data=data[-c(452, 421, 135),])

> summary(m4)

Call:

lm(formula = FACE ~ GENDER + AGE + MARSTAT0 + MARSTAT1 + EDUCATION +

ETHNICITY1 + ETHNICITY2 + ETHNICITY3 + SAGE + SEDUCATION +

NUMHH + INCOME + Sqrincome + CHARITY + Sqrcharity, data = data[-c(452,

421, 135), ])

Residuals:

Min 1Q Median 3Q Max

-1732747 -372655 -136828 72337 8869444

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.102e+06 4.494e+05 -2.452 0.0145 \*

GENDER 8.447e+03 1.553e+05 0.054 0.9566

AGE 1.961e+03 4.678e+03 0.419 0.6753

MARSTAT0 4.877e+05 3.878e+05 1.258 0.2091

MARSTAT1 8.874e+04 1.799e+05 0.493 0.6222

EDUCATION 4.475e+04 1.773e+04 2.524 0.0119 \*

ETHNICITY1 -1.934e+05 1.884e+05 -1.027 0.3050

ETHNICITY2 -2.475e+05 2.161e+05 -1.145 0.2526

ETHNICITY3 -2.065e+05 2.510e+05 -0.823 0.4111

SAGE 6.043e+02 5.888e+03 0.103 0.9183

SEDUCATION 4.262e+04 1.829e+04 2.331 0.0202 \*

NUMHH 7.792e+04 3.330e+04 2.340 0.0197 \*

INCOME 3.477e-01 8.132e-02 4.276 2.3e-05 \*\*\*

Sqrincome -4.644e-09 1.083e-09 -4.286 2.2e-05 \*\*\*

CHARITY -3.424e-01 7.540e-01 -0.454 0.6499

Sqrcharity 2.097e-08 8.298e-08 0.253 0.8006

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 903000 on 481 degrees of freedom

Multiple R-squared: 0.1594, Adjusted R-squared: 0.1332

F-statistic: 6.081 on 15 and 481 DF, p-value: 9.807e-12

Removed the outliers one by one and ran the regression but no use. But one problem got eliminated that quadratic nature of the INCOME and CHARITY variables .

And trying other alternatives.

Removing insignificant variables which are not contributing much to response variable. This decision can be supported by added variable plot.

From the AV plots we can say that below variables are not much contributing to teh response variables

ETHNICITY1,SAGE,ETHNICITY3 and CHARITY

We can try deleting the CHARITY variable first

> m5 <- lm(FACE~GENDER+AGE+MARSTAT0+MARSTAT1+EDUCATION+ETHNICITY1+ETHNICITY2+ETHNICITY3+SAGE+SEDUCATION+NUMHH+INCOME+Sqrincome+Sqrcharity, data=data[-c(452, 421, 135),])

> summary(m5)

Call:

lm(formula = FACE ~ GENDER + AGE + MARSTAT0 + MARSTAT1 + EDUCATION +

ETHNICITY1 + ETHNICITY2 + ETHNICITY3 + SAGE + SEDUCATION +

NUMHH + INCOME + Sqrincome + Sqrcharity, data = data[-c(452,

421, 135), ])

Residuals:

Min 1Q Median 3Q Max

-1689127 -374290 -134431 70643 8869677

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.078e+06 4.459e+05 -2.418 0.0160 \*

GENDER 9.403e+03 1.551e+05 0.061 0.9517

AGE 1.959e+03 4.674e+03 0.419 0.6754

MARSTAT0 4.791e+05 3.870e+05 1.238 0.2163

MARSTAT1 8.880e+04 1.798e+05 0.494 0.6216

EDUCATION 4.396e+04 1.763e+04 2.493 0.0130 \*

ETHNICITY1 -1.992e+05 1.878e+05 -1.061 0.2893

ETHNICITY2 -2.531e+05 2.155e+05 -1.174 0.2409

ETHNICITY3 -2.142e+05 2.502e+05 -0.856 0.3923

SAGE 3.104e+02 5.848e+03 0.053 0.9577

SEDUCATION 4.274e+04 1.827e+04 2.339 0.0197 \*

NUMHH 7.821e+04 3.327e+04 2.351 0.0191 \*

INCOME 3.300e-01 7.133e-02 4.627 4.77e-06 \*\*\*

Sqrincome -4.422e-09 9.664e-10 -4.575 6.05e-06 \*\*\*

Sqrcharity -1.635e-08 1.140e-08 -1.434 0.1523

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 902300 on 482 degrees of freedom

Multiple R-squared: 0.1591, Adjusted R-squared: 0.1346

F-statistic: 6.512 on 14 and 482 DF, p-value: 4.072e-12

Still results are not encouraging.

Tried with introducing new variable squareAGE and ran regression

data$SqrAge <- data$AGE\*data$AGE

> m7 <- lm(FACE~GENDER+AGE+SqrAge+MARSTAT0+MARSTAT1+EDUCATION+ETHNICITY1+ETHNICITY2+ETHNICITY3+SAGE+SEDUCATION+NUMHH+INCOME+Sqrincome, data=data[-c(452, 421, 135),])

> summary(m7)

Call:

lm(formula = FACE ~ GENDER + AGE + SqrAge + MARSTAT0 + MARSTAT1 +

EDUCATION + ETHNICITY1 + ETHNICITY2 + ETHNICITY3 + SAGE +

SEDUCATION + NUMHH + INCOME + Sqrincome, data = data[-c(452,

421, 135), ])

Residuals:

Min 1Q Median 3Q Max

-1629940 -370239 -134190 64410 8853359

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.484e+06 5.655e+05 -2.624 0.00898 \*\*

GENDER 1.273e+04 1.553e+05 0.082 0.93470

AGE 2.309e+04 1.797e+04 1.284 0.19963

SqrAge -2.236e+02 1.841e+02 -1.215 0.22500

MARSTAT0 4.768e+05 3.872e+05 1.231 0.21879

MARSTAT1 6.266e+04 1.814e+05 0.345 0.72995

EDUCATION 4.162e+04 1.774e+04 2.346 0.01941 \*

ETHNICITY1 -1.968e+05 1.880e+05 -1.047 0.29547

ETHNICITY2 -2.606e+05 2.157e+05 -1.208 0.22754

ETHNICITY3 -2.271e+05 2.504e+05 -0.907 0.36500

SAGE 4.296e+02 5.866e+03 0.073 0.94166

SEDUCATION 4.349e+04 1.827e+04 2.380 0.01771 \*

NUMHH 7.366e+04 3.355e+04 2.196 0.02860 \*

INCOME 3.107e-01 7.011e-02 4.431 1.16e-05 \*\*\*

Sqrincome -4.169e-09 9.505e-10 -4.386 1.42e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 902800 on 482 degrees of freedom

Multiple R-squared: 0.158, Adjusted R-squared: 0.1336

F-statistic: 6.463 on 14 and 482 DF, p-value: 5.238e-12

Also trying STEP AIC technique and results are

step <- stepAIC(m7, direction="both")

Start: AIC=13645.76

FACE ~ GENDER + AGE + SqrAge + MARSTAT0 + MARSTAT1 + EDUCATION +

ETHNICITY1 + ETHNICITY2 + ETHNICITY3 + SAGE + SEDUCATION +

NUMHH + INCOME + Sqrincome

Df Sum of Sq RSS AIC

- SAGE 1 4.3705e+09 3.9287e+14 13644

- GENDER 1 5.4771e+09 3.9287e+14 13644

- MARSTAT1 1 9.7231e+10 3.9296e+14 13644

- ETHNICITY3 1 6.7013e+11 3.9353e+14 13645

- ETHNICITY1 1 8.9404e+11 3.9376e+14 13645

- ETHNICITY2 1 1.1899e+12 3.9405e+14 13645

- SqrAge 1 1.2030e+12 3.9407e+14 13645

- MARSTAT0 1 1.2359e+12 3.9410e+14 13645

- AGE 1 1.3445e+12 3.9421e+14 13646

<none> 3.9286e+14 13646

- NUMHH 1 3.9289e+12 3.9679e+14 13649

- EDUCATION 1 4.4841e+12 3.9735e+14 13649

- SEDUCATION 1 4.6162e+12 3.9748e+14 13650

- Sqrincome 1 1.5682e+13 4.0855e+14 13663

- INCOME 1 1.6006e+13 4.0887e+14 13664

Step: AIC=13643.76

FACE ~ GENDER + AGE + SqrAge + MARSTAT0 + MARSTAT1 + EDUCATION +

ETHNICITY1 + ETHNICITY2 + ETHNICITY3 + SEDUCATION + NUMHH +

INCOME + Sqrincome

Df Sum of Sq RSS AIC

- GENDER 1 5.8472e+09 3.9287e+14 13642

- MARSTAT1 1 9.9547e+10 3.9297e+14 13642

- ETHNICITY3 1 6.6603e+11 3.9353e+14 13643

- ETHNICITY1 1 8.8983e+11 3.9376e+14 13643

- ETHNICITY2 1 1.1872e+12 3.9405e+14 13643

- SqrAge 1 1.2006e+12 3.9407e+14 13643

- AGE 1 1.3665e+12 3.9423e+14 13644

<none> 3.9287e+14 13644

- MARSTAT0 1 1.9901e+12 3.9486e+14 13644

+ SAGE 1 4.3705e+09 3.9286e+14 13646

- NUMHH 1 3.9733e+12 3.9684e+14 13647

- EDUCATION 1 4.5511e+12 3.9742e+14 13648

- SEDUCATION 1 4.6126e+12 3.9748e+14 13648

- Sqrincome 1 1.5686e+13 4.0855e+14 13661

- INCOME 1 1.6005e+13 4.0887e+14 13662

Step: AIC=13641.77

FACE ~ AGE + SqrAge + MARSTAT0 + MARSTAT1 + EDUCATION + ETHNICITY1 +

ETHNICITY2 + ETHNICITY3 + SEDUCATION + NUMHH + INCOME + Sqrincome

Df Sum of Sq RSS AIC

- MARSTAT1 1 1.0018e+11 3.9297e+14 13640

- ETHNICITY3 1 6.6600e+11 3.9354e+14 13641

- ETHNICITY1 1 8.9158e+11 3.9376e+14 13641

- SqrAge 1 1.1973e+12 3.9407e+14 13641

- ETHNICITY2 1 1.1975e+12 3.9407e+14 13641

- AGE 1 1.3616e+12 3.9423e+14 13642

<none> 3.9287e+14 13642

- MARSTAT0 1 2.0158e+12 3.9489e+14 13642

+ GENDER 1 5.8472e+09 3.9287e+14 13644

+ SAGE 1 4.7406e+09 3.9287e+14 13644

- NUMHH 1 3.9978e+12 3.9687e+14 13645

- EDUCATION 1 4.5571e+12 3.9743e+14 13646

- SEDUCATION 1 4.8268e+12 3.9770e+14 13646

- Sqrincome 1 1.5701e+13 4.0857e+14 13659

- INCOME 1 1.6010e+13 4.0888e+14 13660

Step: AIC=13639.9

FACE ~ AGE + SqrAge + MARSTAT0 + EDUCATION + ETHNICITY1 + ETHNICITY2 +

ETHNICITY3 + SEDUCATION + NUMHH + INCOME + Sqrincome

Df Sum of Sq RSS AIC

- ETHNICITY3 1 6.7537e+11 3.9365e+14 13639

- ETHNICITY1 1 8.8233e+11 3.9386e+14 13639

- ETHNICITY2 1 1.2348e+12 3.9421e+14 13640

- SqrAge 1 1.3050e+12 3.9428e+14 13640

- AGE 1 1.5170e+12 3.9449e+14 13640

<none> 3.9297e+14 13640

- MARSTAT0 1 2.0450e+12 3.9502e+14 13640

+ MARSTAT1 1 1.0018e+11 3.9287e+14 13642

+ SAGE 1 7.1734e+09 3.9297e+14 13642

+ GENDER 1 6.4789e+09 3.9297e+14 13642

- NUMHH 1 4.1078e+12 3.9708e+14 13643

- EDUCATION 1 4.7236e+12 3.9770e+14 13644

- SEDUCATION 1 5.0322e+12 3.9801e+14 13644

- Sqrincome 1 1.5744e+13 4.0872e+14 13657

- INCOME 1 1.6044e+13 4.0902e+14 13658

Step: AIC=13638.75

FACE ~ AGE + SqrAge + MARSTAT0 + EDUCATION + ETHNICITY1 + ETHNICITY2 +

SEDUCATION + NUMHH + INCOME + Sqrincome

Df Sum of Sq RSS AIC

- ETHNICITY1 1 2.5254e+11 3.9390e+14 13637

- ETHNICITY2 1 5.7207e+11 3.9422e+14 13638

- SqrAge 1 1.2696e+12 3.9492e+14 13638

- AGE 1 1.5082e+12 3.9516e+14 13639

<none> 3.9365e+14 13639

- MARSTAT0 1 2.2246e+12 3.9587e+14 13640

+ ETHNICITY3 1 6.7537e+11 3.9297e+14 13640

+ MARSTAT1 1 1.0955e+11 3.9354e+14 13641

+ GENDER 1 6.4761e+09 3.9364e+14 13641

+ SAGE 1 1.2600e+09 3.9365e+14 13641

- NUMHH 1 3.9094e+12 3.9756e+14 13642

- SEDUCATION 1 5.3139e+12 3.9896e+14 13643

- EDUCATION 1 6.3955e+12 4.0004e+14 13645

- Sqrincome 1 1.5735e+13 4.0938e+14 13656

- INCOME 1 1.6017e+13 4.0967e+14 13657

Step: AIC=13637.07

FACE ~ AGE + SqrAge + MARSTAT0 + EDUCATION + ETHNICITY2 + SEDUCATION +

NUMHH + INCOME + Sqrincome

Df Sum of Sq RSS AIC

- ETHNICITY2 1 3.1953e+11 3.9422e+14 13636

- SqrAge 1 1.3455e+12 3.9525e+14 13637

- AGE 1 1.5768e+12 3.9548e+14 13637

<none> 3.9390e+14 13637

- MARSTAT0 1 2.1175e+12 3.9602e+14 13638

+ ETHNICITY1 1 2.5254e+11 3.9365e+14 13639

+ MARSTAT1 1 9.6201e+10 3.9380e+14 13639

+ ETHNICITY3 1 4.5585e+10 3.9386e+14 13639

+ GENDER 1 7.8299e+09 3.9389e+14 13639

+ SAGE 1 6.3854e+08 3.9390e+14 13639

- NUMHH 1 4.2839e+12 3.9819e+14 13640

- SEDUCATION 1 5.1484e+12 3.9905e+14 13642

- EDUCATION 1 6.1451e+12 4.0005e+14 13643

- Sqrincome 1 1.5772e+13 4.0967e+14 13655

- INCOME 1 1.6050e+13 4.0995e+14 13655

Step: AIC=13635.47

FACE ~ AGE + SqrAge + MARSTAT0 + EDUCATION + SEDUCATION + NUMHH +

INCOME + Sqrincome

Df Sum of Sq RSS AIC

- SqrAge 1 1.2829e+12 3.9550e+14 13635

- AGE 1 1.5233e+12 3.9574e+14 13635

<none> 3.9422e+14 13636

- MARSTAT0 1 2.0849e+12 3.9631e+14 13636

+ ETHNICITY2 1 3.1953e+11 3.9390e+14 13637

+ MARSTAT1 1 1.3431e+11 3.9409e+14 13637

+ GENDER 1 1.6230e+10 3.9420e+14 13638

+ ETHNICITY3 1 1.0416e+10 3.9421e+14 13638

+ SAGE 1 3.2086e+09 3.9422e+14 13638

+ ETHNICITY1 1 5.2649e+06 3.9422e+14 13638

- NUMHH 1 4.2564e+12 3.9848e+14 13639

- SEDUCATION 1 5.3373e+12 3.9956e+14 13640

- EDUCATION 1 6.2631e+12 4.0048e+14 13641

- Sqrincome 1 1.5714e+13 4.0993e+14 13653

- INCOME 1 1.6015e+13 4.1024e+14 13653

Step: AIC=13635.09

FACE ~ AGE + MARSTAT0 + EDUCATION + SEDUCATION + NUMHH + INCOME +

Sqrincome

Df Sum of Sq RSS AIC

- AGE 1 4.6061e+11 3.9596e+14 13634

<none> 3.9550e+14 13635

+ SqrAge 1 1.2829e+12 3.9422e+14 13636

- MARSTAT0 1 2.0758e+12 3.9758e+14 13636

+ ETHNICITY2 1 2.5689e+11 3.9525e+14 13637

+ MARSTAT1 1 2.4531e+11 3.9526e+14 13637

+ GENDER 1 9.8519e+09 3.9549e+14 13637

+ ETHNICITY1 1 8.5974e+09 3.9549e+14 13637

+ SAGE 1 3.0350e+09 3.9550e+14 13637

+ ETHNICITY3 1 2.3809e+09 3.9550e+14 13637

- NUMHH 1 5.2303e+12 4.0073e+14 13640

- SEDUCATION 1 5.3262e+12 4.0083e+14 13640

- EDUCATION 1 6.9055e+12 4.0241e+14 13642

- Sqrincome 1 1.5715e+13 4.1122e+14 13652

- INCOME 1 1.6063e+13 4.1157e+14 13653

Step: AIC=13633.67

FACE ~ MARSTAT0 + EDUCATION + SEDUCATION + NUMHH + INCOME + Sqrincome

Df Sum of Sq RSS AIC

<none> 3.9596e+14 13634

- MARSTAT0 1 1.9393e+12 3.9790e+14 13634

+ AGE 1 4.6061e+11 3.9550e+14 13635

+ MARSTAT1 1 3.8840e+11 3.9558e+14 13635

+ ETHNICITY2 1 2.9838e+11 3.9567e+14 13635

+ SqrAge 1 2.2016e+11 3.9574e+14 13635

+ SAGE 1 2.0825e+11 3.9576e+14 13635

+ ETHNICITY3 1 1.0836e+10 3.9595e+14 13636

+ GENDER 1 1.7706e+09 3.9596e+14 13636

+ ETHNICITY1 1 1.5684e+09 3.9596e+14 13636

- NUMHH 1 4.7837e+12 4.0075e+14 13638

- SEDUCATION 1 5.4272e+12 4.0139e+14 13638

- EDUCATION 1 6.8985e+12 4.0286e+14 13640

- Sqrincome 1 1.6111e+13 4.1208e+14 13652

- INCOME 1 1.6517e+13 4.1248e+14 13652

step$anova

Final Model:

FACE ~ MARSTAT0 + EDUCATION + SEDUCATION + NUMHH + INCOME + Sqrincome

Step Df Deviance Resid. Df Resid. Dev AIC

1 482 3.928628e+14 13645.76

2 - SAGE 1 4.370499e+09 483 3.928671e+14 13643.76

3 - GENDER 1 5.847177e+09 484 3.928730e+14 13641.77

4 - MARSTAT1 1 1.001785e+11 485 3.929732e+14 13639.90

5 - ETHNICITY3 1 6.753690e+11 486 3.936485e+14 13638.75

6 - ETHNICITY1 1 2.525445e+11 487 3.939011e+14 13637.07

7 - ETHNICITY2 1 3.195269e+11 488 3.942206e+14 13635.47

8 - SqrAge 1 1.282880e+12 489 3.955035e+14 13635.09

9 - AGE 1 4.606111e+11 490 3.959641e+14 13633.67