StatLinearModel\_homework3

# Loading the dataset in to the R studio

library(faraway)

## Warning: package 'faraway' was built under R version 3.3.3

data(dvisits)

# structure and summary of the dataset

str(dvisits)

## 'data.frame': 5190 obs. of 19 variables:  
## $ sex : int 1 1 0 0 0 1 1 1 1 0 ...  
## $ age : num 0.19 0.19 0.19 0.19 0.19 0.19 0.19 0.19 0.19 0.19 ...  
## $ agesq : num 0.0361 0.0361 0.0361 0.0361 0.0361 0.0361 0.0361 0.0361 0.0361 0.0361 ...  
## $ income : num 0.55 0.45 0.9 0.15 0.45 0.35 0.55 0.15 0.65 0.15 ...  
## $ levyplus: int 1 1 0 0 0 0 0 0 1 1 ...  
## $ freepoor: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ freerepa: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ illness : int 1 1 3 1 2 5 4 3 2 1 ...  
## $ actdays : int 4 2 0 0 5 1 0 0 0 0 ...  
## $ hscore : int 1 1 0 0 1 9 2 6 5 0 ...  
## $ chcond1 : int 0 0 0 0 1 1 0 0 0 0 ...  
## $ chcond2 : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ doctorco: int 1 1 1 1 1 1 1 1 1 1 ...  
## $ nondocco: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ hospadmi: int 0 0 1 0 0 0 0 0 0 0 ...  
## $ hospdays: int 0 0 4 0 0 0 0 0 0 0 ...  
## $ medicine: int 1 2 2 0 3 1 0 1 1 1 ...  
## $ prescrib: int 1 1 1 0 1 1 0 1 0 1 ...  
## $ nonpresc: int 0 1 1 0 2 0 0 0 1 0 ...

summary(dvisits)

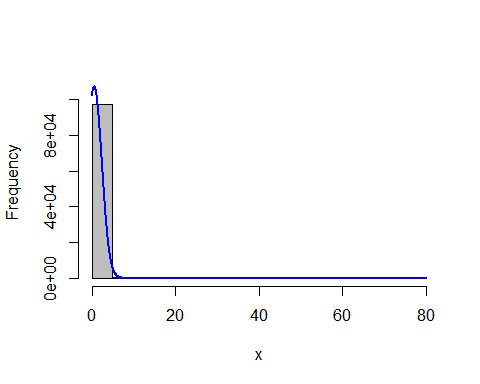
## sex age agesq income   
## Min. :0.0000 Min. :0.1900 Min. :0.0361 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.2200 1st Qu.:0.0484 1st Qu.:0.2500   
## Median :1.0000 Median :0.3200 Median :0.1024 Median :0.5500   
## Mean :0.5206 Mean :0.4064 Mean :0.2071 Mean :0.5832   
## 3rd Qu.:1.0000 3rd Qu.:0.6200 3rd Qu.:0.3844 3rd Qu.:0.9000   
## Max. :1.0000 Max. :0.7200 Max. :0.5184 Max. :1.5000   
## levyplus freepoor freerepa illness   
## Min. :0.0000 Min. :0.00000 Min. :0.0000 Min. :0.000   
## 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.000   
## Median :0.0000 Median :0.00000 Median :0.0000 Median :1.000   
## Mean :0.4428 Mean :0.04277 Mean :0.2102 Mean :1.432   
## 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:2.000   
## Max. :1.0000 Max. :1.00000 Max. :1.0000 Max. :5.000   
## actdays hscore chcond1 chcond2   
## Min. : 0.0000 Min. : 0.000 Min. :0.0000 Min. :0.0000   
## 1st Qu.: 0.0000 1st Qu.: 0.000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median : 0.0000 Median : 0.000 Median :0.0000 Median :0.0000   
## Mean : 0.8619 Mean : 1.218 Mean :0.4031 Mean :0.1166   
## 3rd Qu.: 0.0000 3rd Qu.: 2.000 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :14.0000 Max. :12.000 Max. :1.0000 Max. :1.0000   
## doctorco nondocco hospadmi hospdays   
## Min. :0.0000 Min. : 0.0000 Min. :0.0000 Min. : 0.000   
## 1st Qu.:0.0000 1st Qu.: 0.0000 1st Qu.:0.0000 1st Qu.: 0.000   
## Median :0.0000 Median : 0.0000 Median :0.0000 Median : 0.000   
## Mean :0.3017 Mean : 0.2146 Mean :0.1736 Mean : 1.334   
## 3rd Qu.:0.0000 3rd Qu.: 0.0000 3rd Qu.:0.0000 3rd Qu.: 0.000   
## Max. :9.0000 Max. :11.0000 Max. :5.0000 Max. :80.000   
## medicine prescrib nonpresc   
## Min. :0.000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :1.000 Median :0.0000 Median :0.0000   
## Mean :1.218 Mean :0.8626 Mean :0.3557   
## 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :8.000 Max. :8.0000 Max. :8.0000

# looking at the data distribution

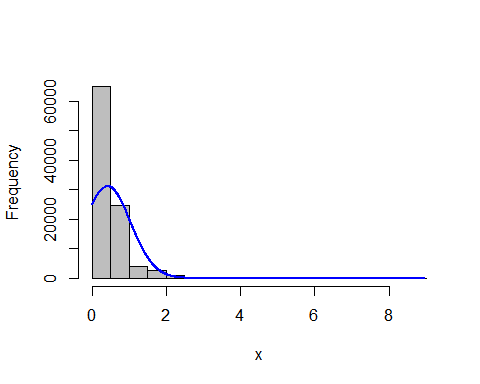
library(rcompanion)

## Warning: package 'rcompanion' was built under R version 3.3.3

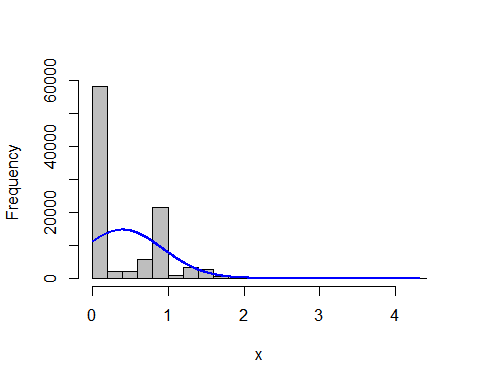
plotNormalHistogram(dvisits)

 #Looks like the distribution of the data is right Skewed. Performing the regression with the skewed data may cause problems with the analysis. The results might not be accurate. So we will apply the common transformations for right-skewed data: square root, cube root, and log. The square root transformation improves the distribution of the data.

#taking square root transformation  
dvisit\_sqrt = sqrt(dvisits)  
plotNormalHistogram(dvisit\_sqrt)



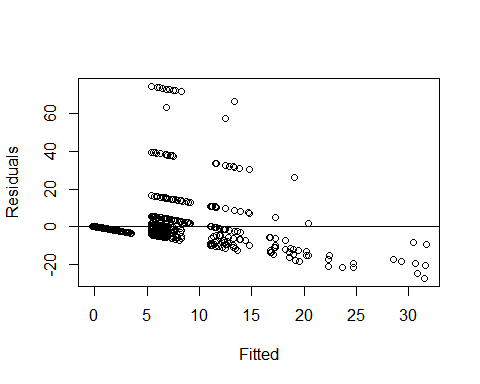
#taking cube root transformation  
dvisits\_cub = sign(dvisits) \* abs(dvisits)^(1/3)  
plotNormalHistogram(dvisits\_cub)

 #The cube root transformation is stronger than the square root transformation.Here we are trying to transform the entire dataset as it involves multiple attributes. This also fix the non constant variance issue. The dvisist dataset when modelled with the right skewed distribution produced the non constant variance.Below is the plot for the non constant variance of the input data.

skewed\_model <- lm(hospdays ~ freerepa + actdays + hospadmi+ chcond2 +age , data = dvisits)  
summary(skewed\_model)

##   
## Call:  
## lm(formula = hospdays ~ freerepa + actdays + hospadmi + chcond2 +   
## age, data = dvisits)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -27.516 -1.072 -0.058 0.226 74.584   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.47674 0.17559 -2.715 0.006648 \*\*   
## freerepa 0.84337 0.22683 3.718 0.000203 \*\*\*  
## actdays 0.10647 0.02669 3.988 6.75e-05 \*\*\*  
## hospadmi 5.64290 0.15071 37.442 < 2e-16 \*\*\*  
## chcond2 0.85288 0.23816 3.581 0.000345 \*\*\*  
## age 1.13775 0.44943 2.532 0.011385 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.269 on 5184 degrees of freedom  
## Multiple R-squared: 0.2596, Adjusted R-squared: 0.2589   
## F-statistic: 363.5 on 5 and 5184 DF, p-value: < 2.2e-16

plot(fitted(skewed\_model), residuals(skewed\_model), xlab = "Fitted", ylab = "Residuals")  
abline(h = 0)

 #The above figure shows that there is no constant variance and our model also does not give a good R squared value. That is why, we are taking the cube root transformation of the dataset to achieve a better model.

# Now let us perform the data cleaning and preprocessing

#checking for missing values  
sum(is.na(dvisits\_cub))

## [1] 0

#checking for multicollinearity  
summary(dvisits\_cub)

## sex age agesq income   
## Min. :0.0000 Min. :0.5749 Min. :0.3305 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.6037 1st Qu.:0.3644 1st Qu.:0.6300   
## Median :1.0000 Median :0.6840 Median :0.4678 Median :0.8193   
## Mean :0.5206 Mean :0.7197 Mean :0.5333 Mean :0.7895   
## 3rd Qu.:1.0000 3rd Qu.:0.8527 3rd Qu.:0.7271 3rd Qu.:0.9655   
## Max. :1.0000 Max. :0.8963 Max. :0.8033 Max. :1.1447   
## levyplus freepoor freerepa illness   
## Min. :0.0000 Min. :0.00000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.00000 Median :0.0000 Median :1.0000   
## Mean :0.4428 Mean :0.04277 Mean :0.2102 Mean :0.8574   
## 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:1.2599   
## Max. :1.0000 Max. :1.00000 Max. :1.0000 Max. :1.7100   
## actdays hscore chcond1 chcond2   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
## Mean :0.2352 Mean :0.5587 Mean :0.4031 Mean :0.1166   
## 3rd Qu.:0.0000 3rd Qu.:1.2599 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :2.4101 Max. :2.2894 Max. :1.0000 Max. :1.0000   
## doctorco nondocco hospadmi hospdays   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
## Mean :0.2225 Mean :0.1125 Mean :0.1436 Mean :0.2505   
## 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000   
## Max. :2.0801 Max. :2.2240 Max. :1.7100 Max. :4.3089   
## medicine prescrib nonpresc   
## Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :1.0000 Median :0.0000 Median :0.0000   
## Mean :0.7022 Mean :0.4987 Mean :0.2859   
## 3rd Qu.:1.2599 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :2.0000 Max. :2.0000 Max. :2.0000

vif(dvisits\_cub)

## sex age agesq income levyplus freepoor   
## 1.234971 540.061932 550.282645 1.407108 1.553946 1.162277   
## freerepa illness actdays hscore chcond1 chcond2   
## 2.456229 1.455059 1.307031 1.225407 1.453788 1.418848   
## doctorco nondocco hospadmi hospdays medicine prescrib   
## 1.306995 1.090627 6.295817 6.386564 12.812947 9.941837   
## nonpresc   
## 4.656563

?dvisits

## starting httpd help server ... done

# Checking for the multicollinearity using the Variance Inflation Factor has given the above result. We have eliminate the varibles which are highly correlated. VIF value of greater than 10 said to be sufferring from multicollinearity.

#removing the highly collinear features  
  
dvisit\_newdata <- dvisits\_cub[, -3]  
dvisit\_newdata <- dvisit\_newdata[, -16]  
summary(dvisit\_newdata)

## sex age income levyplus   
## Min. :0.0000 Min. :0.5749 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.6037 1st Qu.:0.6300 1st Qu.:0.0000   
## Median :1.0000 Median :0.6840 Median :0.8193 Median :0.0000   
## Mean :0.5206 Mean :0.7197 Mean :0.7895 Mean :0.4428   
## 3rd Qu.:1.0000 3rd Qu.:0.8527 3rd Qu.:0.9655 3rd Qu.:1.0000   
## Max. :1.0000 Max. :0.8963 Max. :1.1447 Max. :1.0000   
## freepoor freerepa illness actdays   
## Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.00000 Median :0.0000 Median :1.0000 Median :0.0000   
## Mean :0.04277 Mean :0.2102 Mean :0.8574 Mean :0.2352   
## 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:1.2599 3rd Qu.:0.0000   
## Max. :1.00000 Max. :1.0000 Max. :1.7100 Max. :2.4101   
## hscore chcond1 chcond2 doctorco   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
## Mean :0.5587 Mean :0.4031 Mean :0.1166 Mean :0.2225   
## 3rd Qu.:1.2599 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:0.0000   
## Max. :2.2894 Max. :1.0000 Max. :1.0000 Max. :2.0801   
## nondocco hospadmi hospdays prescrib   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
## Mean :0.1125 Mean :0.1436 Mean :0.2505 Mean :0.4987   
## 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:1.0000   
## Max. :2.2240 Max. :1.7100 Max. :4.3089 Max. :2.0000   
## nonpresc   
## Min. :0.0000   
## 1st Qu.:0.0000   
## Median :0.0000   
## Mean :0.2859   
## 3rd Qu.:1.0000   
## Max. :2.0000

# After removing the highly correlated variables, check for the mulicollinearity again to ensure all the variables are in the lower VIF range.

vif(dvisit\_newdata)

## sex age income levyplus freepoor freerepa illness actdays   
## 1.232791 1.925995 1.268796 1.551955 1.161911 2.417777 1.448244 1.306008   
## hscore chcond1 chcond2 doctorco nondocco hospadmi hospdays prescrib   
## 1.223825 1.452327 1.393652 1.306990 1.089295 6.294368 6.376720 1.659035   
## nonpresc   
## 1.062481

# In order to perform the regression, we will consider only the variables which are signficant and that has p value less than 0.05. To achieve this we can do varible selection.

# Variable selection involves three methods - Forward, Backward and Both. We have to consider the model with the lesser AIC value.

#Perform the variable selection method to sustain the significant variables in the model  
#nullmodel  
nullmodel <- lm(hospdays ~ 1, data = dvisit\_newdata)  
summary(nullmodel)

##   
## Call:  
## lm(formula = hospdays ~ 1, data = dvisit\_newdata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.2505 -0.2505 -0.2505 -0.2505 4.0584   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.250462 0.009578 26.15 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.69 on 5189 degrees of freedom

#fullmodel  
fullmodel <- lm(hospdays~., data = dvisit\_newdata)  
summary(fullmodel)

##   
## Call:  
## lm(formula = hospdays ~ ., data = dvisit\_newdata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.63033 -0.03926 0.00841 0.02881 2.64193   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.0950448 0.0339012 -2.804 0.005073 \*\*   
## sex -0.0173484 0.0084391 -2.056 0.039860 \*   
## age 0.1416030 0.0425007 3.332 0.000869 \*\*\*  
## income -0.0105872 0.0210343 -0.503 0.614752   
## levyplus -0.0084137 0.0095264 -0.883 0.377175   
## freepoor -0.0113336 0.0202350 -0.560 0.575439   
## freerepa 0.0307287 0.0144899 2.121 0.033995 \*   
## illness -0.0028866 0.0076944 -0.375 0.707560   
## actdays 0.0176827 0.0070674 2.502 0.012380 \*   
## hscore -0.0035116 0.0060247 -0.583 0.560006   
## chcond1 0.0043535 0.0093325 0.466 0.640890   
## chcond2 0.0309505 0.0139676 2.216 0.026743 \*   
## doctorco 0.0004064 0.0096171 0.042 0.966299   
## nondocco 0.0555309 0.0106994 5.190 2.18e-07 \*\*\*  
## hospadmi 1.6930156 0.0107892 156.917 < 2e-16 \*\*\*  
## prescrib 0.0111653 0.0078116 1.429 0.152976   
## nonpresc -0.0050802 0.0081108 -0.626 0.531113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2737 on 5173 degrees of freedom  
## Multiple R-squared: 0.8432, Adjusted R-squared: 0.8427   
## F-statistic: 1738 on 16 and 5173 DF, p-value: < 2.2e-16

#variable selection - forward  
forward <- step(nullmodel, scope=list(lower=nullmodel, upper=fullmodel), direction="forward")

## Start: AIC=-3850.8  
## hospdays ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + hospadmi 1 2072.41 398.0 -13324.1  
## + actdays 1 113.13 2357.3 -4092.1  
## + doctorco 1 82.99 2387.4 -4026.1  
## + chcond2 1 80.93 2389.5 -4021.7  
## + prescrib 1 80.67 2389.7 -4021.1  
## + nondocco 1 59.52 2410.9 -3975.4  
## + illness 1 47.61 2422.8 -3949.8  
## + hscore 1 43.41 2427.0 -3940.8  
## + freerepa 1 23.73 2446.7 -3898.9  
## + age 1 22.93 2447.5 -3897.2  
## + income 1 13.02 2457.4 -3876.2  
## + sex 1 2.98 2467.4 -3855.1  
## + chcond1 1 2.74 2467.7 -3854.6  
## + levyplus 1 1.38 2469.0 -3851.7  
## <none> 2470.4 -3850.8  
## + nonpresc 1 0.58 2469.8 -3850.0  
## + freepoor 1 0.06 2470.3 -3848.9  
##   
## Step: AIC=-13324.08  
## hospdays ~ hospadmi  
##   
## Df Sum of Sq RSS AIC  
## + age 1 5.0547 392.95 -13388  
## + freerepa 1 4.6404 393.36 -13383  
## + nondocco 1 3.8267 394.18 -13372  
## + prescrib 1 2.7375 395.27 -13358  
## + chcond2 1 1.7890 396.21 -13346  
## + actdays 1 1.4535 396.55 -13341  
## + doctorco 1 0.8047 397.20 -13333  
## + levyplus 1 0.6882 397.31 -13331  
## + income 1 0.6374 397.37 -13330  
## + illness 1 0.5830 397.42 -13330  
## + chcond1 1 0.2790 397.72 -13326  
## + freepoor 1 0.2181 397.78 -13325  
## <none> 398.00 -13324  
## + hscore 1 0.1223 397.88 -13324  
## + sex 1 0.0817 397.92 -13323  
## + nonpresc 1 0.0725 397.93 -13323  
##   
## Step: AIC=-13388.41  
## hospdays ~ hospadmi + age  
##   
## Df Sum of Sq RSS AIC  
## + nondocco 1 2.76665 390.18 -13423  
## + actdays 1 1.27714 391.67 -13403  
## + chcond2 1 1.14870 391.80 -13402  
## + freerepa 1 1.06342 391.88 -13400  
## + prescrib 1 0.62887 392.32 -13395  
## + levyplus 1 0.54337 392.40 -13394  
## + doctorco 1 0.33814 392.61 -13391  
## + illness 1 0.19943 392.75 -13389  
## <none> 392.95 -13388  
## + income 1 0.12467 392.82 -13388  
## + hscore 1 0.11537 392.83 -13388  
## + sex 1 0.11066 392.84 -13388  
## + nonpresc 1 0.03358 392.91 -13387  
## + chcond1 1 0.01838 392.93 -13387  
## + freepoor 1 0.00807 392.94 -13386  
##   
## Step: AIC=-13423.08  
## hospdays ~ hospadmi + age + nondocco  
##   
## Df Sum of Sq RSS AIC  
## + freerepa 1 0.96845 389.21 -13434  
## + chcond2 1 0.78781 389.39 -13432  
## + actdays 1 0.75846 389.42 -13431  
## + levyplus 1 0.56464 389.62 -13429  
## + prescrib 1 0.38054 389.80 -13426  
## + doctorco 1 0.19128 389.99 -13424  
## + sex 1 0.17574 390.01 -13423  
## <none> 390.18 -13423  
## + income 1 0.09942 390.08 -13422  
## + illness 1 0.07465 390.11 -13422  
## + nonpresc 1 0.05365 390.13 -13422  
## + hscore 1 0.02494 390.16 -13421  
## + chcond1 1 0.02255 390.16 -13421  
## + freepoor 1 0.00588 390.18 -13421  
##   
## Step: AIC=-13433.98  
## hospdays ~ hospadmi + age + nondocco + freerepa  
##   
## Df Sum of Sq RSS AIC  
## + actdays 1 0.74237 388.47 -13442  
## + chcond2 1 0.62846 388.58 -13440  
## + prescrib 1 0.27121 388.94 -13436  
## + sex 1 0.26673 388.95 -13436  
## <none> 389.21 -13434  
## + doctorco 1 0.14388 389.07 -13434  
## + levyplus 1 0.06867 389.14 -13433  
## + nonpresc 1 0.03805 389.17 -13432  
## + illness 1 0.03256 389.18 -13432  
## + chcond1 1 0.01861 389.19 -13432  
## + hscore 1 0.00898 389.20 -13432  
## + freepoor 1 0.00395 389.21 -13432  
## + income 1 0.00105 389.21 -13432  
##   
## Step: AIC=-13441.89  
## hospdays ~ hospadmi + age + nondocco + freerepa + actdays  
##   
## Df Sum of Sq RSS AIC  
## + chcond2 1 0.44077 388.03 -13446  
## + sex 1 0.28440 388.19 -13444  
## <none> 388.47 -13442  
## + prescrib 1 0.13610 388.33 -13442  
## + levyplus 1 0.06607 388.40 -13441  
## + nonpresc 1 0.05939 388.41 -13441  
## + chcond1 1 0.01578 388.45 -13440  
## + hscore 1 0.01132 388.46 -13440  
## + doctorco 1 0.00774 388.46 -13440  
## + freepoor 1 0.00427 388.47 -13440  
## + income 1 0.00086 388.47 -13440  
## + illness 1 0.00038 388.47 -13440  
##   
## Step: AIC=-13445.78  
## hospdays ~ hospadmi + age + nondocco + freerepa + actdays + chcond2  
##   
## Df Sum of Sq RSS AIC  
## + sex 1 0.265499 387.76 -13447  
## <none> 388.03 -13446  
## + prescrib 1 0.078288 387.95 -13445  
## + levyplus 1 0.072201 387.96 -13445  
## + nonpresc 1 0.065793 387.96 -13445  
## + hscore 1 0.032238 388.00 -13444  
## + chcond1 1 0.016254 388.01 -13444  
## + illness 1 0.008291 388.02 -13444  
## + freepoor 1 0.006286 388.02 -13444  
## + doctorco 1 0.004315 388.03 -13444  
## + income 1 0.001052 388.03 -13444  
##   
## Step: AIC=-13447.33  
## hospdays ~ hospadmi + age + nondocco + freerepa + actdays + chcond2 +   
## sex  
##   
## Df Sum of Sq RSS AIC  
## + prescrib 1 0.174309 387.59 -13448  
## <none> 387.76 -13447  
## + nonpresc 1 0.049842 387.71 -13446  
## + levyplus 1 0.038172 387.73 -13446  
## + chcond1 1 0.028275 387.74 -13446  
## + hscore 1 0.026668 387.74 -13446  
## + income 1 0.015817 387.75 -13446  
## + doctorco 1 0.008799 387.76 -13446  
## + freepoor 1 0.007481 387.76 -13445  
## + illness 1 0.002417 387.76 -13445  
##   
## Step: AIC=-13447.67  
## hospdays ~ hospadmi + age + nondocco + freerepa + actdays + chcond2 +   
## sex + prescrib  
##   
## Df Sum of Sq RSS AIC  
## <none> 387.59 -13448  
## + levyplus 1 0.048135 387.54 -13446  
## + hscore 1 0.040416 387.55 -13446  
## + nonpresc 1 0.039738 387.55 -13446  
## + illness 1 0.023810 387.57 -13446  
## + income 1 0.014484 387.58 -13446  
## + freepoor 1 0.005880 387.58 -13446  
## + chcond1 1 0.004700 387.59 -13446  
## + doctorco 1 0.000026 387.59 -13446

#variable selection - backward  
backward <- step(fullmodel, data=dvisit\_newdata, direction="backward")

## Start: AIC=-13434.07  
## hospdays ~ sex + age + income + levyplus + freepoor + freerepa +   
## illness + actdays + hscore + chcond1 + chcond2 + doctorco +   
## nondocco + hospadmi + prescrib + nonpresc  
##   
## Df Sum of Sq RSS AIC  
## - doctorco 1 0.00 387.41 -13436.1  
## - illness 1 0.01 387.42 -13435.9  
## - chcond1 1 0.02 387.43 -13435.9  
## - income 1 0.02 387.43 -13435.8  
## - freepoor 1 0.02 387.43 -13435.8  
## - hscore 1 0.03 387.44 -13435.7  
## - nonpresc 1 0.03 387.44 -13435.7  
## - levyplus 1 0.06 387.47 -13435.3  
## <none> 387.41 -13434.1  
## - prescrib 1 0.15 387.56 -13434.0  
## - sex 1 0.32 387.73 -13431.8  
## - freerepa 1 0.34 387.75 -13431.6  
## - chcond2 1 0.37 387.78 -13431.1  
## - actdays 1 0.47 387.88 -13429.8  
## - age 1 0.83 388.24 -13424.9  
## - nondocco 1 2.02 389.43 -13409.1  
## - hospadmi 1 1844.04 2231.45 -4348.8  
##   
## Step: AIC=-13436.07  
## hospdays ~ sex + age + income + levyplus + freepoor + freerepa +   
## illness + actdays + hscore + chcond1 + chcond2 + nondocco +   
## hospadmi + prescrib + nonpresc  
##   
## Df Sum of Sq RSS AIC  
## - illness 1 0.01 387.42 -13437.9  
## - chcond1 1 0.02 387.43 -13437.9  
## - income 1 0.02 387.43 -13437.8  
## - freepoor 1 0.02 387.43 -13437.8  
## - hscore 1 0.03 387.44 -13437.7  
## - nonpresc 1 0.03 387.44 -13437.7  
## - levyplus 1 0.06 387.47 -13437.3  
## <none> 387.41 -13436.1  
## - prescrib 1 0.16 387.57 -13435.9  
## - sex 1 0.32 387.73 -13433.8  
## - freerepa 1 0.34 387.75 -13433.5  
## - chcond2 1 0.37 387.78 -13433.1  
## - actdays 1 0.51 387.92 -13431.2  
## - age 1 0.83 388.24 -13426.9  
## - nondocco 1 2.02 389.43 -13411.1  
## - hospadmi 1 1854.41 2241.82 -4326.7  
##   
## Step: AIC=-13437.93  
## hospdays ~ sex + age + income + levyplus + freepoor + freerepa +   
## actdays + hscore + chcond1 + chcond2 + nondocco + hospadmi +   
## prescrib + nonpresc  
##   
## Df Sum of Sq RSS AIC  
## - chcond1 1 0.01 387.43 -13439.8  
## - income 1 0.02 387.44 -13439.7  
## - freepoor 1 0.02 387.44 -13439.6  
## - nonpresc 1 0.04 387.46 -13439.4  
## - hscore 1 0.04 387.46 -13439.4  
## - levyplus 1 0.06 387.48 -13439.1  
## <none> 387.42 -13437.9  
## - prescrib 1 0.15 387.57 -13437.9  
## - sex 1 0.32 387.74 -13435.7  
## - freerepa 1 0.33 387.75 -13435.5  
## - chcond2 1 0.36 387.78 -13435.1  
## - actdays 1 0.50 387.93 -13433.2  
## - age 1 0.85 388.27 -13428.6  
## - nondocco 1 2.01 389.43 -13413.0  
## - hospadmi 1 1854.86 2242.28 -4327.7  
##   
## Step: AIC=-13439.78  
## hospdays ~ sex + age + income + levyplus + freepoor + freerepa +   
## actdays + hscore + chcond2 + nondocco + hospadmi + prescrib +   
## nonpresc  
##   
## Df Sum of Sq RSS AIC  
## - income 1 0.02 387.45 -13441.5  
## - freepoor 1 0.02 387.46 -13441.5  
## - nonpresc 1 0.03 387.47 -13441.3  
## - hscore 1 0.03 387.47 -13441.3  
## - levyplus 1 0.06 387.49 -13441.0  
## <none> 387.43 -13439.8  
## - prescrib 1 0.18 387.62 -13439.3  
## - sex 1 0.32 387.75 -13437.6  
## - freerepa 1 0.34 387.77 -13437.3  
## - chcond2 1 0.37 387.80 -13436.8  
## - actdays 1 0.50 387.94 -13435.0  
## - age 1 0.91 388.35 -13429.5  
## - nondocco 1 2.02 389.45 -13414.8  
## - hospadmi 1 1857.35 2244.79 -4323.9  
##   
## Step: AIC=-13441.53  
## hospdays ~ sex + age + levyplus + freepoor + freerepa + actdays +   
## hscore + chcond2 + nondocco + hospadmi + prescrib + nonpresc  
##   
## Df Sum of Sq RSS AIC  
## - freepoor 1 0.02 387.47 -13443.3  
## - hscore 1 0.03 387.48 -13443.1  
## - nonpresc 1 0.03 387.48 -13443.1  
## - levyplus 1 0.06 387.51 -13442.7  
## <none> 387.45 -13441.5  
## - prescrib 1 0.19 387.64 -13441.0  
## - sex 1 0.30 387.75 -13439.5  
## - chcond2 1 0.37 387.82 -13438.6  
## - freerepa 1 0.38 387.83 -13438.4  
## - actdays 1 0.50 387.95 -13436.8  
## - age 1 0.92 388.37 -13431.3  
## - nondocco 1 2.02 389.47 -13416.5  
## - hospadmi 1 1859.55 2247.00 -4320.7  
##   
## Step: AIC=-13443.33  
## hospdays ~ sex + age + levyplus + freerepa + actdays + hscore +   
## chcond2 + nondocco + hospadmi + prescrib + nonpresc  
##   
## Df Sum of Sq RSS AIC  
## - nonpresc 1 0.03 387.50 -13445  
## - hscore 1 0.04 387.50 -13445  
## - levyplus 1 0.05 387.52 -13445  
## <none> 387.47 -13443  
## - prescrib 1 0.19 387.65 -13443  
## - sex 1 0.30 387.77 -13441  
## - chcond2 1 0.36 387.83 -13440  
## - freerepa 1 0.42 387.88 -13440  
## - actdays 1 0.50 387.97 -13439  
## - age 1 0.93 388.40 -13433  
## - nondocco 1 2.03 389.49 -13418  
## - hospadmi 1 1860.27 2247.73 -4321  
##   
## Step: AIC=-13444.88  
## hospdays ~ sex + age + levyplus + freerepa + actdays + hscore +   
## chcond2 + nondocco + hospadmi + prescrib  
##   
## Df Sum of Sq RSS AIC  
## - hscore 1 0.04 387.54 -13446.3  
## - levyplus 1 0.05 387.55 -13446.2  
## <none> 387.50 -13444.9  
## - prescrib 1 0.20 387.70 -13444.2  
## - sex 1 0.32 387.82 -13442.6  
## - chcond2 1 0.36 387.86 -13442.1  
## - freerepa 1 0.43 387.93 -13441.2  
## - actdays 1 0.49 387.99 -13440.3  
## - age 1 0.93 388.43 -13434.5  
## - nondocco 1 2.02 389.52 -13419.9  
## - hospadmi 1 1861.74 2249.24 -4319.6  
##   
## Step: AIC=-13446.31  
## hospdays ~ sex + age + levyplus + freerepa + actdays + chcond2 +   
## nondocco + hospadmi + prescrib  
##   
## Df Sum of Sq RSS AIC  
## - levyplus 1 0.05 387.59 -13447.7  
## <none> 387.54 -13446.3  
## - prescrib 1 0.18 387.73 -13445.8  
## - sex 1 0.32 387.86 -13444.0  
## - chcond2 1 0.34 387.88 -13443.8  
## - freerepa 1 0.42 387.96 -13442.7  
## - actdays 1 0.46 388.00 -13442.2  
## - age 1 0.96 388.50 -13435.5  
## - nondocco 1 1.99 389.53 -13421.7  
## - hospadmi 1 1867.12 2254.66 -4309.1  
##   
## Step: AIC=-13447.67  
## hospdays ~ sex + age + freerepa + actdays + chcond2 + nondocco +   
## hospadmi + prescrib  
##   
## Df Sum of Sq RSS AIC  
## <none> 387.59 -13447.7  
## - prescrib 1 0.17 387.76 -13447.3  
## - chcond2 1 0.34 387.93 -13445.2  
## - sex 1 0.36 387.95 -13444.8  
## - actdays 1 0.46 388.05 -13443.5  
## - freerepa 1 0.85 388.44 -13438.3  
## - age 1 0.92 388.51 -13437.4  
## - nondocco 1 1.98 389.57 -13423.2  
## - hospadmi 1 1867.07 2254.66 -4311.1

#variable selection - both directions  
stepwise <- step(nullmodel, scope = list(upper=fullmodel), data=dvisit\_newdata, direction="both")

## Start: AIC=-3850.8  
## hospdays ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + hospadmi 1 2072.41 398.0 -13324.1  
## + actdays 1 113.13 2357.3 -4092.1  
## + doctorco 1 82.99 2387.4 -4026.1  
## + chcond2 1 80.93 2389.5 -4021.7  
## + prescrib 1 80.67 2389.7 -4021.1  
## + nondocco 1 59.52 2410.9 -3975.4  
## + illness 1 47.61 2422.8 -3949.8  
## + hscore 1 43.41 2427.0 -3940.8  
## + freerepa 1 23.73 2446.7 -3898.9  
## + age 1 22.93 2447.5 -3897.2  
## + income 1 13.02 2457.4 -3876.2  
## + sex 1 2.98 2467.4 -3855.1  
## + chcond1 1 2.74 2467.7 -3854.6  
## + levyplus 1 1.38 2469.0 -3851.7  
## <none> 2470.4 -3850.8  
## + nonpresc 1 0.58 2469.8 -3850.0  
## + freepoor 1 0.06 2470.3 -3848.9  
##   
## Step: AIC=-13324.08  
## hospdays ~ hospadmi  
##   
## Df Sum of Sq RSS AIC  
## + age 1 5.05 392.95 -13388.4  
## + freerepa 1 4.64 393.36 -13382.9  
## + nondocco 1 3.83 394.18 -13372.2  
## + prescrib 1 2.74 395.27 -13357.9  
## + chcond2 1 1.79 396.21 -13345.5  
## + actdays 1 1.45 396.55 -13341.1  
## + doctorco 1 0.80 397.20 -13332.6  
## + levyplus 1 0.69 397.31 -13331.1  
## + income 1 0.64 397.37 -13330.4  
## + illness 1 0.58 397.42 -13329.7  
## + chcond1 1 0.28 397.72 -13325.7  
## + freepoor 1 0.22 397.78 -13324.9  
## <none> 398.00 -13324.1  
## + hscore 1 0.12 397.88 -13323.7  
## + sex 1 0.08 397.92 -13323.1  
## + nonpresc 1 0.07 397.93 -13323.0  
## - hospadmi 1 2072.41 2470.41 -3850.8  
##   
## Step: AIC=-13388.41  
## hospdays ~ hospadmi + age  
##   
## Df Sum of Sq RSS AIC  
## + nondocco 1 2.77 390.18 -13423.1  
## + actdays 1 1.28 391.67 -13403.3  
## + chcond2 1 1.15 391.80 -13401.6  
## + freerepa 1 1.06 391.88 -13400.5  
## + prescrib 1 0.63 392.32 -13394.7  
## + levyplus 1 0.54 392.40 -13393.6  
## + doctorco 1 0.34 392.61 -13390.9  
## + illness 1 0.20 392.75 -13389.0  
## <none> 392.95 -13388.4  
## + income 1 0.12 392.82 -13388.1  
## + hscore 1 0.12 392.83 -13387.9  
## + sex 1 0.11 392.84 -13387.9  
## + nonpresc 1 0.03 392.91 -13386.9  
## + chcond1 1 0.02 392.93 -13386.7  
## + freepoor 1 0.01 392.94 -13386.5  
## - age 1 5.05 398.00 -13324.1  
## - hospadmi 1 2054.53 2447.48 -3897.2  
##   
## Step: AIC=-13423.08  
## hospdays ~ hospadmi + age + nondocco  
##   
## Df Sum of Sq RSS AIC  
## + freerepa 1 0.97 389.21 -13434.0  
## + chcond2 1 0.79 389.39 -13431.6  
## + actdays 1 0.76 389.42 -13431.2  
## + levyplus 1 0.56 389.62 -13428.6  
## + prescrib 1 0.38 389.80 -13426.1  
## + doctorco 1 0.19 389.99 -13423.6  
## + sex 1 0.18 390.01 -13423.4  
## <none> 390.18 -13423.1  
## + income 1 0.10 390.08 -13422.4  
## + illness 1 0.07 390.11 -13422.1  
## + nonpresc 1 0.05 390.13 -13421.8  
## + hscore 1 0.02 390.16 -13421.4  
## + chcond1 1 0.02 390.16 -13421.4  
## + freepoor 1 0.01 390.18 -13421.2  
## - nondocco 1 2.77 392.95 -13388.4  
## - age 1 3.99 394.18 -13372.2  
## - hospadmi 1 2006.84 2397.03 -4003.3  
##   
## Step: AIC=-13433.98  
## hospdays ~ hospadmi + age + nondocco + freerepa  
##   
## Df Sum of Sq RSS AIC  
## + actdays 1 0.74 388.47 -13441.9  
## + chcond2 1 0.63 388.58 -13440.4  
## + prescrib 1 0.27 388.94 -13435.6  
## + sex 1 0.27 388.95 -13435.5  
## <none> 389.21 -13434.0  
## + doctorco 1 0.14 389.07 -13433.9  
## + levyplus 1 0.07 389.14 -13432.9  
## + nonpresc 1 0.04 389.17 -13432.5  
## + illness 1 0.03 389.18 -13432.4  
## + chcond1 1 0.02 389.19 -13432.2  
## + hscore 1 0.01 389.20 -13432.1  
## + freepoor 1 0.00 389.21 -13432.0  
## + income 1 0.00 389.21 -13432.0  
## - freerepa 1 0.97 390.18 -13423.1  
## - age 1 1.11 390.33 -13421.2  
## - nondocco 1 2.67 391.88 -13400.5  
## - hospadmi 1 2002.46 2391.68 -4012.9  
##   
## Step: AIC=-13441.89  
## hospdays ~ hospadmi + age + nondocco + freerepa + actdays  
##   
## Df Sum of Sq RSS AIC  
## + chcond2 1 0.44 388.03 -13445.8  
## + sex 1 0.28 388.19 -13443.7  
## <none> 388.47 -13441.9  
## + prescrib 1 0.14 388.33 -13441.7  
## + levyplus 1 0.07 388.40 -13440.8  
## + nonpresc 1 0.06 388.41 -13440.7  
## + chcond1 1 0.02 388.45 -13440.1  
## + hscore 1 0.01 388.46 -13440.0  
## + doctorco 1 0.01 388.46 -13440.0  
## + freepoor 1 0.00 388.47 -13439.9  
## + income 1 0.00 388.47 -13439.9  
## + illness 1 0.00 388.47 -13439.9  
## - actdays 1 0.74 389.21 -13434.0  
## - freerepa 1 0.95 389.42 -13431.2  
## - age 1 1.11 389.58 -13429.1  
## - nondocco 1 2.17 390.64 -13415.0  
## - hospadmi 1 1917.52 2305.99 -4200.2  
##   
## Step: AIC=-13445.78  
## hospdays ~ hospadmi + age + nondocco + freerepa + actdays + chcond2  
##   
## Df Sum of Sq RSS AIC  
## + sex 1 0.27 387.76 -13447.3  
## <none> 388.03 -13445.8  
## + prescrib 1 0.08 387.95 -13444.8  
## + levyplus 1 0.07 387.96 -13444.7  
## + nonpresc 1 0.07 387.96 -13444.7  
## + hscore 1 0.03 388.00 -13444.2  
## + chcond1 1 0.02 388.01 -13444.0  
## + illness 1 0.01 388.02 -13443.9  
## + freepoor 1 0.01 388.02 -13443.9  
## + doctorco 1 0.00 388.03 -13443.8  
## + income 1 0.00 388.03 -13443.8  
## - chcond2 1 0.44 388.47 -13441.9  
## - actdays 1 0.55 388.58 -13440.4  
## - freerepa 1 0.82 388.85 -13436.9  
## - age 1 1.07 389.10 -13433.5  
## - nondocco 1 1.99 390.02 -13421.3  
## - hospadmi 1 1884.47 2272.50 -4274.2  
##   
## Step: AIC=-13447.33  
## hospdays ~ hospadmi + age + nondocco + freerepa + actdays + chcond2 +   
## sex  
##   
## Df Sum of Sq RSS AIC  
## + prescrib 1 0.17 387.59 -13447.7  
## <none> 387.76 -13447.3  
## + nonpresc 1 0.05 387.71 -13446.0  
## + levyplus 1 0.04 387.73 -13445.8  
## - sex 1 0.27 388.03 -13445.8  
## + chcond1 1 0.03 387.74 -13445.7  
## + hscore 1 0.03 387.74 -13445.7  
## + income 1 0.02 387.75 -13445.5  
## + doctorco 1 0.01 387.76 -13445.5  
## + freepoor 1 0.01 387.76 -13445.4  
## + illness 1 0.00 387.76 -13445.4  
## - chcond2 1 0.42 388.19 -13443.7  
## - actdays 1 0.57 388.34 -13441.7  
## - freerepa 1 0.90 388.67 -13437.2  
## - age 1 1.22 388.98 -13433.1  
## - nondocco 1 2.05 389.82 -13421.9  
## - hospadmi 1 1884.70 2272.46 -4272.3  
##   
## Step: AIC=-13447.67  
## hospdays ~ hospadmi + age + nondocco + freerepa + actdays + chcond2 +   
## sex + prescrib  
##   
## Df Sum of Sq RSS AIC  
## <none> 387.59 -13447.7  
## - prescrib 1 0.17 387.76 -13447.3  
## + levyplus 1 0.05 387.54 -13446.3  
## + hscore 1 0.04 387.55 -13446.2  
## + nonpresc 1 0.04 387.55 -13446.2  
## + illness 1 0.02 387.57 -13446.0  
## + income 1 0.01 387.58 -13445.9  
## + freepoor 1 0.01 387.58 -13445.7  
## + chcond1 1 0.00 387.59 -13445.7  
## + doctorco 1 0.00 387.59 -13445.7  
## - chcond2 1 0.34 387.93 -13445.2  
## - sex 1 0.36 387.95 -13444.8  
## - actdays 1 0.46 388.05 -13443.5  
## - freerepa 1 0.85 388.44 -13438.3  
## - age 1 0.92 388.51 -13437.4  
## - nondocco 1 1.98 389.57 -13423.2  
## - hospadmi 1 1867.07 2254.66 -4311.1

# The lowest AIC value in all the 3 methods turns out to be same(13447.67) and now we will build our regression model with the above variables.

#model with lowest AIC value  
try\_model1 <- lm(hospdays ~ freerepa + actdays + hospadmi+ chcond2 +age+ nondocco+sex+prescrib , data = dvisit\_newdata)  
summary(try\_model1)

##   
## Call:  
## lm(formula = hospdays ~ freerepa + actdays + hospadmi + chcond2 +   
## age + nondocco + sex + prescrib, data = dvisit\_newdata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.63872 -0.03920 0.00704 0.02747 2.64659   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.110846 0.026894 -4.122 3.82e-05 \*\*\*  
## freerepa 0.039334 0.011672 3.370 0.000757 \*\*\*  
## actdays 0.016358 0.006583 2.485 0.012984 \*   
## hospadmi 1.692837 0.010716 157.980 < 2e-16 \*\*\*  
## chcond2 0.026587 0.012538 2.121 0.034003 \*   
## age 0.139803 0.039972 3.498 0.000474 \*\*\*  
## nondocco 0.054921 0.010668 5.148 2.73e-07 \*\*\*  
## sex -0.018012 0.008193 -2.198 0.027972 \*   
## prescrib 0.011034 0.007228 1.526 0.126961   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2735 on 5181 degrees of freedom  
## Multiple R-squared: 0.8431, Adjusted R-squared: 0.8429   
## F-statistic: 3480 on 8 and 5181 DF, p-value: < 2.2e-16

# Again remove the insignificant varibles and build the model

try\_model2 <- lm(hospdays ~ freerepa + actdays + hospadmi+ chcond2 +age+ nondocco+sex , data = dvisit\_newdata)  
summary(try\_model2)

##   
## Call:  
## lm(formula = hospdays ~ freerepa + actdays + hospadmi + chcond2 +   
## age + nondocco + sex, data = dvisit\_newdata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.64441 -0.04212 0.00785 0.02810 2.65520   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.119577 0.026282 -4.550 5.49e-06 \*\*\*  
## freerepa 0.040487 0.011649 3.476 0.000514 \*\*\*  
## actdays 0.017974 0.006498 2.766 0.005691 \*\*   
## hospadmi 1.694270 0.010676 158.703 < 2e-16 \*\*\*  
## chcond2 0.029440 0.012399 2.374 0.017614 \*   
## age 0.155629 0.038609 4.031 5.64e-05 \*\*\*  
## nondocco 0.055813 0.010653 5.239 1.68e-07 \*\*\*  
## sex -0.014973 0.007949 -1.884 0.059670 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2735 on 5182 degrees of freedom  
## Multiple R-squared: 0.843, Adjusted R-squared: 0.8428   
## F-statistic: 3976 on 7 and 5182 DF, p-value: < 2.2e-16

finalmodel <- lm(hospdays ~ freerepa + actdays + hospadmi+ chcond2 +age+ nondocco , data = dvisit\_newdata)  
summary(finalmodel)

##   
## Call:  
## lm(formula = hospdays ~ freerepa + actdays + hospadmi + chcond2 +   
## age + nondocco, data = dvisit\_newdata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.65364 -0.04307 0.00648 0.03147 2.64622   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.118241 0.026279 -4.499 6.96e-06 \*\*\*  
## freerepa 0.038327 0.011595 3.306 0.000954 \*\*\*  
## actdays 0.017686 0.006498 2.722 0.006512 \*\*   
## hospadmi 1.694121 0.010678 158.655 < 2e-16 \*\*\*  
## chcond2 0.030081 0.012397 2.426 0.015283 \*   
## age 0.143741 0.038099 3.773 0.000163 \*\*\*  
## nondocco 0.054858 0.010644 5.154 2.64e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2736 on 5183 degrees of freedom  
## Multiple R-squared: 0.8429, Adjusted R-squared: 0.8427   
## F-statistic: 4636 on 6 and 5183 DF, p-value: < 2.2e-16

# This is the final regression model for predicting the hospdays with respect to other variables in the given dataset.

# Q>no:2 - Why is your model a good/reasonable model? Check the constant variance assumption for the errors.

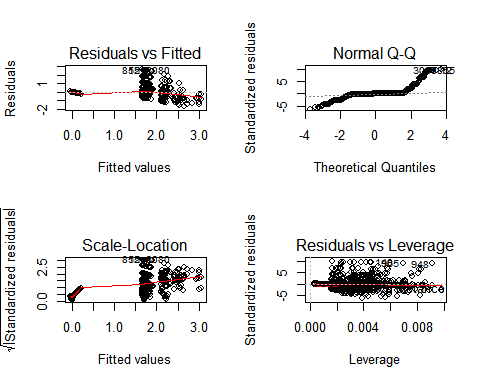
# This can be considered as the better model for the following reasons:

# *all the variables are significant with the p value less than 0.05. #* The adjusted R squared value is 0.8427 ie., 84.2%

# \*The standard residual error is 0.2736 which is minimal when compared to all the other model.

# Constant Variance Assumption for the eror with the regression diagnostics

par(mfrow = c(2, 2))  
plot(finalmodel)



plot(fitted(finalmodel), residuals(finalmodel), xlab = "Fitted", ylab = "Residuals")  
abline(h = 0)

