STA302 final project

```
dataframe<-read.csv("FP_dataset.csv",header=T)
dataframe_sel<-dataframe[,-c(1:3)]
set.seed(1004712965)
sample<-sample.int(n=1508,size=1131,replace=F)
train_set0<-dataframe_sel[sample,]
test_set0<-dataframe_sel[-sample,]
```

```
model_test<-lm(ADM_RATE ~ .,data=train_set0)
summary(model_test)
```

```
##
## Call:
## lm(formula = ADM_RATE ~ ., data = train_set0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76063 -0.10258  0.00525  0.11543  0.43040
##
## Coefficients: (1 not defined because of singularities)
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        3.991e-01  3.905e-01   1.022 0.306957
## STABBRAL          -6.347e-02  1.598e-01  -0.397 0.691338
## STABBRAR          -1.195e-01  1.548e-01  -0.772 0.440219
## STABBRAZ           2.625e-02  1.524e-01   0.172 0.863306
## STABBRCA          -6.999e-02  1.399e-01  -0.500 0.616991
## STABBRCO          -5.584e-02  1.481e-01  -0.377 0.706175
## STABBRCT           6.304e-02  1.506e-01   0.418 0.675694
## STABBRDC          -1.328e-01  1.714e-01  -0.774 0.438819
## STABBRDE          -1.285e-01  1.762e-01  -0.729 0.465928
## STABBRFL          -1.195e-01  1.543e-01  -0.774 0.439090
## STABBRFM          -1.292e-01  2.412e-01  -0.536 0.592259
## STABBRGA          -1.976e-01  1.576e-01  -1.254 0.209986
## STABBRHI           1.548e-01  1.705e-01   0.908 0.364129
## STABBRIA          -8.715e-02  1.507e-01  -0.578 0.563096
## STABBRID           7.890e-03  1.604e-01   0.049 0.960770
## STABBRIL          -1.273e-01  1.457e-01  -0.874 0.382282
## STABBRIN          -3.964e-02  1.493e-01  -0.265 0.790679
## STABBRKS          -1.506e-01  1.517e-01  -0.992 0.321352
## STABBRKY          -5.939e-02  1.581e-01  -0.376 0.707298
## STABBRLA          -1.074e-01  1.591e-01  -0.675 0.499943
## STABBRMA          -1.009e-02  1.483e-01  -0.068 0.945763
## STABBRMD          -2.239e-02  1.554e-01  -0.144 0.885466
## STABBRME          -7.603e-02  1.563e-01  -0.486 0.626868
## STABBRMI          -8.820e-02  1.465e-01  -0.602 0.547179
## STABBRMN          -1.557e-01  1.480e-01  -1.052 0.293048
## STABBRMO          -1.318e-01  1.496e-01  -0.881 0.378412
## STABBRMS          -4.789e-02  1.661e-01  -0.288 0.773147
```

```
## STABBRMT          -3.966e-03  1.568e-01  -0.025 0.979822
## STABBRNC          -1.868e-01  1.520e-01  -1.229 0.219261
## STABBRND          -1.928e-01  1.584e-01  -1.218 0.223605
## STABBRNE          -1.509e-01  1.592e-01  -0.948 0.343184
## STABBRNH          -1.282e-01  1.613e-01  -0.795 0.426954
## STABBRNJ           1.457e-02  1.484e-01   0.098 0.921827
## STABBRNM          -2.253e-01  1.716e-01  -1.313 0.189562
## STABBRNV           9.686e-03  1.839e-01   0.053 0.957994
## STABBRNY           1.343e-02  1.458e-01   0.092 0.926637
## STABBROH          -9.900e-02  1.500e-01  -0.660 0.509324
## STABBROK          -9.059e-02  1.501e-01  -0.604 0.546176
## STABBROR           7.333e-02  1.417e-01   0.517 0.604918
## STABBRPA          -1.511e-02  1.476e-01  -0.102 0.918473
## STABBRPR          -3.265e-01  2.283e-01  -1.430 0.152954
## STABBRRI           4.605e-02  1.596e-01   0.289 0.772963
## STABBRSC          -1.730e-01  1.596e-01  -1.084 0.278636
## STABBRSD          -7.751e-02  1.592e-01  -0.487 0.626514
## STABBRTN          -9.405e-02  1.535e-01  -0.613 0.540318
## STABBRTX          -1.498e-01  1.497e-01  -1.000 0.317325
## STABBRUT           9.158e-03  1.605e-01   0.057 0.954506
## STABBRVA          -3.956e-02  1.541e-01  -0.257 0.797467
## STABBRVI           2.419e-01  2.226e-01   1.087 0.277382
## STABBRVT          -2.393e-02  1.545e-01  -0.155 0.876947
## STABBRWA           5.051e-02  1.408e-01   0.359 0.719853
## STABBRWI          -1.380e-01  1.468e-01  -0.941 0.347120
## STABBRWV          -3.978e-02  1.561e-01  -0.255 0.798889
## NUMBRANCH          5.640e-03  1.929e-03   2.924 0.003527 **
## CONTROL           -3.867e-02  1.816e-02  -2.129 0.033452 *
## REGION                    NA         NA      NA       NA
## HBCU              -5.629e-02  4.454e-02  -1.264 0.206532
## PBI                1.999e-02  4.994e-02   0.400 0.689090
## TRIBAL             4.067e-01  1.655e-01   2.457 0.014178 *
## HSI                5.627e-02  2.661e-02   2.115 0.034680 *
## WOMENONLY          9.043e-02  6.472e-02   1.397 0.162663
## COSTT4_A          -1.884e-06  7.455e-07  -2.528 0.011624 *
## AVGFACSAL         -2.787e-05  3.621e-06  -7.698 3.17e-14 ***
## PFTFAC            -5.315e-02  2.413e-02  -2.202 0.027865 *
## PCTPELL           -4.464e-02  7.249e-02  -0.616 0.538196
## UG25ABV           -3.548e-02  5.325e-02  -0.666 0.505309
## INC_PCT_LO         4.186e-02  1.412e-01   0.297 0.766876
## PAR_ED_PCT_1STGEN  2.016e-02  1.282e-01   0.157 0.875035
## FEMALE             1.276e-01  4.787e-02   2.665 0.007815 **
## MD_FAMINC          6.068e-07  7.424e-07   0.817 0.413857
## PCT_WHITE          3.989e-03  3.265e-03   1.222 0.222025
## PCT_BLACK          4.391e-03  3.530e-03   1.244 0.213708
## PCT_ASIAN          1.919e-04  5.674e-03   0.034 0.973032
## PCT_HISPANIC       3.158e-03  1.864e-03   1.695 0.090452 .
## PCT_BA             2.067e-02  5.577e-03   3.706 0.000221 ***
## PCT_GRAD_PROF     -3.159e-02  7.351e-03  -4.297 1.89e-05 ***
## PCT_BORN_US        2.654e-03  2.029e-03   1.308 0.191148
## POVERTY_RATE      -3.402e-03  4.967e-03  -0.685 0.493531
## UNEMP_RATE        -8.485e-03  2.300e-02  -0.369 0.712316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.1701 on 1053 degrees of freedom
## Multiple R-squared:  0.3107, Adjusted R-squared:  0.2603
## F-statistic: 6.164 on 77 and 1053 DF,  p-value: < 2.2e-16
```

```
dataframe_sel2<-dataframe_sel[,-4]
train_set<-dataframe_sel2[sample,]
test_set<-dataframe_sel2[-sample,]
```

```
library(MASS)
stepAIC(lm(ADM_RATE ~ 1, data=train_set),
        scope=list(upper=lm(ADM_RATE ~ ., data = train_set)),
        direction = "forward", k=2)
```

```
stepAIC(lm(ADM_RATE ~ ., data=train_set),
        scope=list(lower=lm(ADM_RATE ~ 1, data = train_set)),
        direction = "backward", k=2)
```

```
stepAIC(lm(ADM_RATE ~ 1, data=train_set),
        scope=list(upper=lm(ADM_RATE ~ ., data = train_set)),
        direction = "forward", k=log(nrow(train_set)))
```

```
stepAIC(lm(ADM_RATE ~ ., data=train_set),
        scope=list(lower=lm(ADM_RATE ~ 1, data = train_set)),
        direction = "backward", k=log(nrow(train_set)))
```

```
stepAIC(lm(ADM_RATE ~ ., data=train_set), direction = "both", k=2)
```

```
stepAIC(lm(ADM_RATE ~ ., data=train_set), direction = "both", k=log(nrow(train_set)))
```

```
model1<-lm(ADM_RATE ~ AVGFACSAL + CONTROL + POVERTY_RATE +
    PCT_BLACK + NUMBRANCH + FEMALE + COSTT4_A + PFTFAC + MD_FAMINC +
    PCT_GRAD_PROF + UNEMP_RATE + HSI + PCT_BORN_US, data = train_set)
summary(model1)
```

```
model2<-lm(ADM_RATE ~ NUMBRANCH + CONTROL + TRIBAL + HSI +
    COSTT4_A + AVGFACSAL + PFTFAC + PAR_ED_PCT_1STGEN + FEMALE +
    MD_FAMINC + PCT_WHITE + PCT_BA + PCT_GRAD_PROF + POVERTY_RATE +
    UNEMP_RATE, data = train_set)
summary(model2)
```

```
##
## Call:
## lm(formula = ADM_RATE ~ NUMBRANCH + CONTROL + TRIBAL + HSI +
##     COSTT4_A + AVGFACSAL + PFTFAC + PAR_ED_PCT_1STGEN + FEMALE +
##     MD_FAMINC + PCT_WHITE + PCT_BA + PCT_GRAD_PROF + POVERTY_RATE +
##     UNEMP_RATE, data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7576 -0.1202  0.0086  0.1277  0.4263
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         5.985e-01  1.071e-01   5.590 2.85e-08 ***
## NUMBRANCH           6.893e-03  1.787e-03   3.858 0.000121 ***
## CONTROL            -4.155e-02  1.608e-02  -2.584 0.009889 **
```

```
## TRIBAL             1.896e-01  1.239e-01   1.531 0.126077
## HSI                3.637e-02  2.089e-02   1.741 0.081950 .
## COSTT4_A          -2.013e-06  6.820e-07  -2.951 0.003229 **
## AVGFACSAL         -2.771e-05  3.200e-06  -8.659  < 2e-16 ***
## PFTFAC            -5.465e-02  2.258e-02  -2.420 0.015670 *
## PAR_ED_PCT_1STGEN  1.542e-01  1.013e-01   1.522 0.128221
## FEMALE             1.512e-01  4.423e-02   3.418 0.000654 ***
## MD_FAMINC          1.251e-06  4.843e-07   2.584 0.009898 **
## PCT_WHITE          2.385e-03  5.325e-04   4.478 8.30e-06 ***
## PCT_BA             7.133e-03  3.172e-03   2.248 0.024744 *
## PCT_GRAD_PROF     -9.679e-03  4.091e-03  -2.366 0.018161 *
## POVERTY_RATE      -7.786e-03  2.107e-03  -3.695 0.000231 ***
## UNEMP_RATE         3.206e-02  1.363e-02   2.352 0.018854 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1728 on 1115 degrees of freedom
## Multiple R-squared:  0.2469, Adjusted R-squared:  0.2368
## F-statistic: 24.37 on 15 and 1115 DF,  p-value: < 2.2e-16
```

```
model3<-lm(ADM_RATE ~ AVGFACSAL + CONTROL + POVERTY_RATE +
    PCT_BLACK + NUMBRANCH + FEMALE + COSTT4_A, data = train_set)
summary(model3)
```

```
##
## Call:
## lm(formula = ADM_RATE ~ AVGFACSAL + CONTROL + POVERTY_RATE +
##     PCT_BLACK + NUMBRANCH + FEMALE + COSTT4_A, data = train_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76899 -0.12250  0.01219  0.13212  0.40592
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.026e+00  4.247e-02  24.173  < 2e-16 ***
## AVGFACSAL    -3.015e-05  2.817e-06 -10.702  < 2e-16 ***
## CONTROL      -4.515e-02  1.523e-02  -2.964 0.003100 **
## POVERTY_RATE -5.003e-03  9.047e-04  -5.530 3.99e-08 ***
## PCT_BLACK    -2.538e-03  4.995e-04  -5.082 4.37e-07 ***
## NUMBRANCH     6.971e-03  1.785e-03   3.906 9.94e-05 ***
## FEMALE        1.674e-01  4.351e-02   3.846 0.000127 ***
## COSTT4_A     -2.033e-06  5.573e-07  -3.648 0.000277 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1743 on 1123 degrees of freedom
## Multiple R-squared:  0.2278, Adjusted R-squared:  0.223
## F-statistic: 47.32 on 7 and 1123 DF,  p-value: < 2.2e-16
```

```
model4<-lm(ADM_RATE ~ NUMBRANCH + HSI + COSTT4_A + AVGFACSAL +
    FEMALE + PCT_WHITE + POVERTY_RATE, data = train_set)
summary(model4)
```

```
##
```

```
## Call:
## lm(formula = ADM_RATE ~ NUMBRANCH + HSI + COSTT4_A + AVGFACSAL +
##     FEMALE + PCT_WHITE + POVERTY_RATE, data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7946 -0.1228  0.0107  0.1321  0.4003
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.497e-01  5.607e-02  13.371  < 2e-16 ***
## NUMBRANCH     7.290e-03  1.783e-03   4.088 4.66e-05 ***
## HSI           5.861e-02  1.942e-02   3.019 0.002596 **
## COSTT4_A     -3.135e-06  3.813e-07  -8.222 5.48e-16 ***
## AVGFACSAL    -2.412e-05  2.378e-06 -10.143  < 2e-16 ***
## FEMALE        1.554e-01  4.374e-02   3.553 0.000396 ***
## PCT_WHITE     2.211e-03  4.346e-04   5.088 4.23e-07 ***
## POVERTY_RATE -6.126e-03  1.019e-03  -6.013 2.46e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1745 on 1123 degrees of freedom
## Multiple R-squared:  0.2265, Adjusted R-squared:  0.2217
## F-statistic: 46.98 on 7 and 1123 DF,  p-value: < 2.2e-16
```

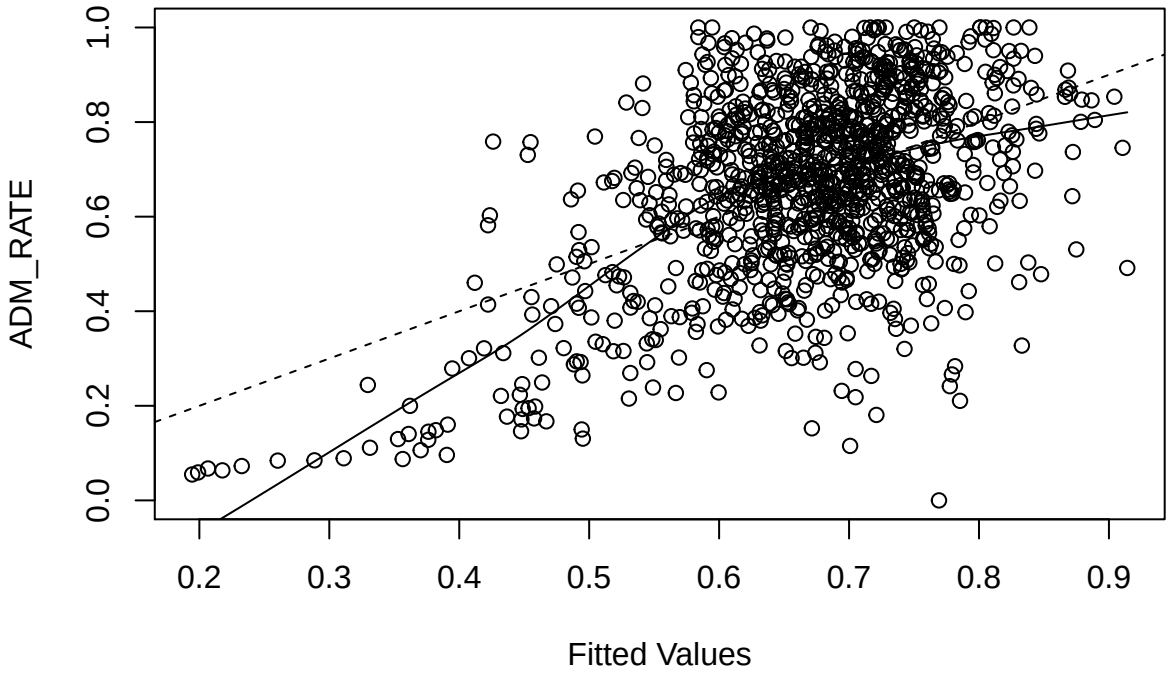
```
model5<-lm(ADM_RATE ~ NUMBRANCH +HSI + COSTT4_A + AVGFACSAL +
    FEMALE + PCT_WHITE + POVERTY_RATE, data = train_set)
summary(model5)
```

```
##
## Call:
## lm(formula = ADM_RATE ~ NUMBRANCH + HSI + COSTT4_A + AVGFACSAL +
##     FEMALE + PCT_WHITE + POVERTY_RATE, data = train_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7946 -0.1228  0.0107  0.1321  0.4003
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.497e-01  5.607e-02  13.371  < 2e-16 ***
## NUMBRANCH     7.290e-03  1.783e-03   4.088 4.66e-05 ***
## HSI           5.861e-02  1.942e-02   3.019 0.002596 **
## COSTT4_A     -3.135e-06  3.813e-07  -8.222 5.48e-16 ***
## AVGFACSAL    -2.412e-05  2.378e-06 -10.143  < 2e-16 ***
## FEMALE        1.554e-01  4.374e-02   3.553 0.000396 ***
## PCT_WHITE     2.211e-03  4.346e-04   5.088 4.23e-07 ***
## POVERTY_RATE -6.126e-03  1.019e-03  -6.013 2.46e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1745 on 1123 degrees of freedom
## Multiple R-squared:  0.2265, Adjusted R-squared:  0.2217
## F-statistic: 46.98 on 7 and 1123 DF,  p-value: < 2.2e-16
```

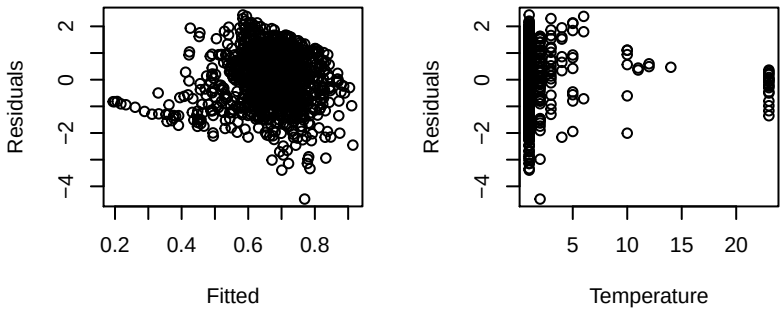
```
select_criteria = function(model, n)
{
  SSres <- sum(model$residuals^2)
  Rsq_adj <- summary(model)$adj.r.squared
  p <- length(model$coefficients) - 1
  AIC <- n*log(SSres/n) + 2*p
  AICc <- AIC + (2*(p+2)*(p+3)/(n-p-1))
  BIC <- n*log(SSres/n) + (p+2)*log(n)
  res <- c(SSres, Rsq_adj, AIC, AICc, BIC,p)
  names(res) <- c("SSres", "Rsq_adj", "AIC", "AIC_c", "BIC","p")
  return(res)
}
n<-nrow(train_set)
results<-rbind(select_criteria(model1,n),select_criteria(model2,n),select_criteria(model3,n),select_cri
results
```

```
##         SSres   Rsq_adj       AIC     AIC_c       BIC  p
## [1,] 33.48444 0.2336582 -3954.867 -3954.438 -3875.405 13
## [2,] 33.28747 0.2367996 -3957.540 -3956.991 -3868.015 15
## [3,] 34.13455 0.2229534 -3945.119 -3944.959 -3895.842  7
## [4,] 34.18971 0.2216978 -3943.293 -3943.133 -3894.015  7
```

```
plot(train_set$ADM_RATE ~ model1$fitted.values, xlab="Fitted Values", ylab="ADM_RATE")
abline(a = 0, b = 1, lty=2)
lines(lowess(model1$fitted.values, train_set$ADM_RATE))
```




```
par(mfrow=c(2,3))
plot(rstandard(model1) ~ fitted(model1), xlab="Fitted", ylab="Residuals")
plot(rstandard(model1) ~ train_set$NUMBRANCH, xlab="Temperature", ylab="Residuals")
```




```
install.packages("car")
library(car)
```

```
install.packages("tidyverse")
library(tidyverse)
```

```
new_data<-train_set%>%
    filter(ADM_RATE>0)
multi_mod <- lm(cbind(new_data$ADM_RATE, new_data$AVGFACSAL,new_data$CONTROL,new_data$POVERTY_RATE,new_
bc <- powerTransform(multi_mod)
summary(bc)
```

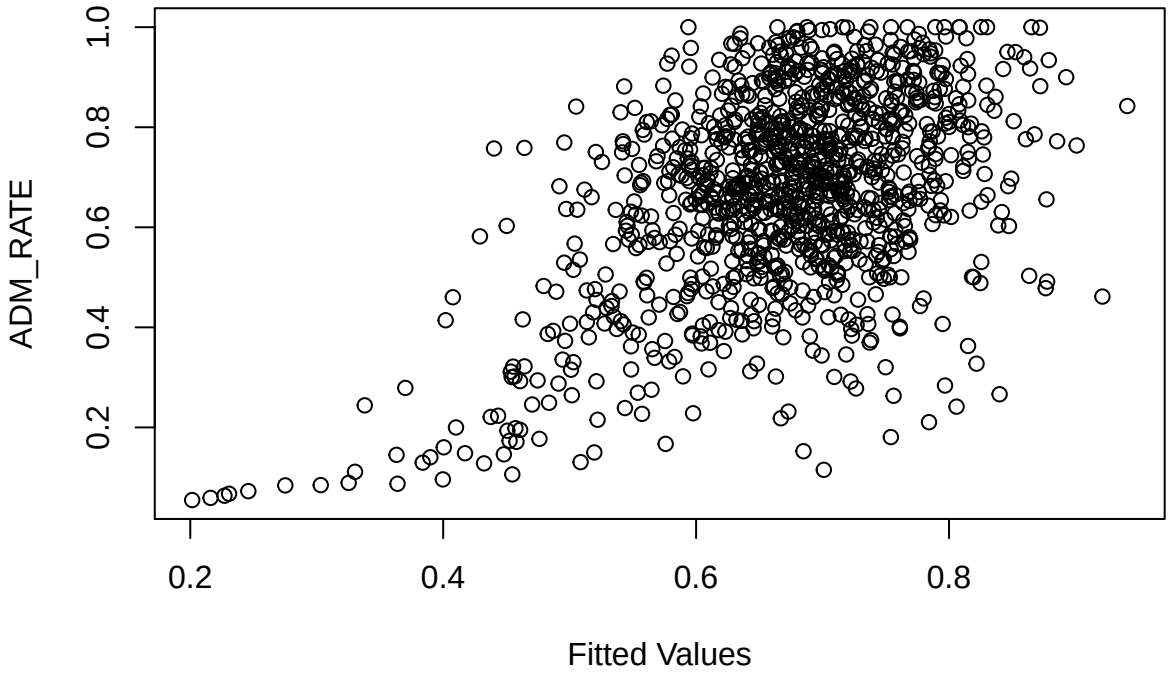
```
## bcPower Transformations to Multinormality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1    1.4564        1.46       1.3054       1.6074
## Y2    0.2318        0.33       0.1299       0.3336
## Y3    0.0882        0.00      -0.0864       0.2628
## Y4   -0.8492       -0.85      -0.9509      -0.7475
## Y5    0.1137        0.11       0.0609       0.1664
## Y6   -6.3000       -6.30      -6.6913      -5.9087
## Y7    1.4595        1.46       1.3149       1.6041
## Y8    0.2999        0.33       0.2007       0.3992
##
## Likelihood ratio test that transformation parameters are equal to 0
##  (all log transformations)
##                                           LRT df       pval
## LR test, lambda = (0 0 0 0 0 0 0 0) 4782.49  8 < 2.22e-16
##
## Likelihood ratio test that no transformations are needed
##                                             LRT df       pval
## LR test, lambda = (1 1 1 1 1 1 1 1) 10160.61  8 < 2.22e-16
```

```
modelt<-lm(ADM_RATE~ AVGFACSAL + I(log(CONTROL)) + POVERTY_RATE +
    I(log(PCT_BLACK)) + I(NUMBRANCH^-6) + I(FEMALE^1.5) + COSTT4_A, data = new_data)
summary(modelt)
```

```
##
## Call:
## lm(formula = ADM_RATE ~ AVGFACSAL + I(log(CONTROL)) + POVERTY_RATE +
##     I(log(PCT_BLACK)) + I(NUMBRANCH^-6) + I(FEMALE^1.5) + COSTT4_A,
##     data = new_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58581 -0.11921  0.01114  0.12876  0.40606
```

```
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.123e+00  3.505e-02  32.048  < 2e-16 ***
## AVGFACSAL          -3.100e-05  2.876e-06 -10.778  < 2e-16 ***
## I(log(CONTROL))    -8.699e-02  2.478e-02  -3.510 0.000465 ***
## POVERTY_RATE       -5.120e-03  8.949e-04  -5.721 1.35e-08 ***
## I(log(PCT_BLACK)) -3.127e-02  5.858e-03  -5.338 1.14e-07 ***
## I(NUMBRANCH^-6)    -8.297e-02  1.658e-02  -5.005 6.48e-07 ***
## I(FEMALE^1.5)       1.586e-01  3.870e-02   4.100 4.44e-05 ***
## COSTT4_A           -1.429e-06  5.973e-07  -2.393 0.016894 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1721 on 1122 degrees of freedom
## Multiple R-squared:  0.2407, Adjusted R-squared:  0.236
## F-statistic: 50.82 on 7 and 1122 DF,  p-value: < 2.2e-16
```

```
plot(new_data$ADM_RATE ~ modelt$fitted.values, xlab="Fitted Values", ylab="ADM_RATE")
```




```
model_3c<-lm(ADM_RATE ~ AVGFACSAL + CONTROL + POVERTY_RATE +
I(NUMBRANCH^-6) + PCT_BLACK + FEMALE + COSTT4_A , data = train_set)
summary(model_3c)
```

```
##
## Call:
## lm(formula = ADM_RATE ~ AVGFACSAL + CONTROL + POVERTY_RATE +
##     I(NUMBRANCH^-6) + PCT_BLACK + FEMALE + COSTT4_A, data = train_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82212 -0.11961  0.01339  0.13025  0.41735
##
```

```
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.122e+00  4.445e-02  25.231  < 2e-16 ***
## AVGFACSAL       -3.135e-05  2.824e-06 -11.104  < 2e-16 ***
## CONTROL         -5.329e-02  1.516e-02  -3.515 0.000458 ***
## POVERTY_RATE    -5.247e-03  9.031e-04  -5.810 8.13e-09 ***
## I(NUMBRANCH^-6) -7.162e-02  1.677e-02  -4.271 2.11e-05 ***
## PCT_BLACK       -2.462e-03  4.998e-04  -4.927 9.61e-07 ***
## FEMALE           1.623e-01  4.342e-02   3.738 0.000195 ***
## COSTT4_A        -1.831e-06  5.594e-07  -3.272 0.001100 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1741 on 1123 degrees of freedom
## Multiple R-squared:  0.2298, Adjusted R-squared:  0.225
## F-statistic: 47.86 on 7 and 1123 DF,  p-value: < 2.2e-16
```

```
h <- hatvalues(model_3c)
threshold <- 2 * (length(model_3c$coefficients)/nrow(train_set))

w <- which(h > threshold)
length(w)
```

```
## [1] 119
```

```
r <- rstandard(model_3c)
which(r>=4|r<=-4)
```

```
## 829
## 207
```

```
D <- cooks.distance(model_3c)
cutoff <- qf(0.5, 8, 754-8, lower.tail=T)
which(D > cutoff)
```

```
## named integer(0)
```

```
fits <- dffits(model_3c)
cutoff <- 2*sqrt(8/754)
which(abs(fits) > cutoff)
```

```
##  179   79  352 1491  759 1439  412  272  454  829  571 1449 1504  992 1426  812
##   48   72   78   83  120  143  165  183  189  207  222  238  242  251  254  300
##  741  912  217 1118 1492 1454 1501 1424   93  625 1427  106 1489  460 1447  134
##  315  348  382  393  405  434  441  448  471  502  617  633  703  729  740  746
##  563  633  438 1432  508 1470  955  118 1452  804  929  866    4  794 1455  949
##  791  830  834  852  858  863  878  885  905  948  969  983  987  990  995 1005
## 1312   62  249 1450  476
## 1008 1021 1028 1078 1098
```

```
dfb <- dfbetas(model_3c)
cutoff <- 2/sqrt(754)
which(abs(dfb[,2]) > cutoff)
```

```
##   79  531  352 1491 1488  759  829   76 1110  808  306  992 1351  812  273  912
##   72   74   78   83   92  120  207  221  227  230  246  251  268  300  337  348
## 1493  217  898 1118 1492 1501  702   81   93 1121 1495 1301  106   82   78  726
##  375  382  385  393  405  441  444  454  471  491  501  555  633  639  659  663
```

```
##  376  233 1489  134  129  563  438  508  955  756 1233  929 1268  866 1120  949
##  672  681  703  746  762  791  834  858  878  918  942  969  978  983 1002 1005
##   62  580  877  177  799  168  409
## 1021 1026 1035 1059 1080 1091 1119
```

```
train_set1<-train_set[-c(746,740,721,633,300,242,207,83,441,56,174,471,405,703),]
```

```
model_f<-lm(ADM_RATE ~ AVGFACSAL + CONTROL + POVERTY_RATE +
I(NUMBRANCH^-6) + PCT_BLACK + FEMALE + COSTT4_A, data = train_set1)
summary(model_f)
```

```
##
## Call:
## lm(formula = ADM_RATE ~ AVGFACSAL + CONTROL + POVERTY_RATE +
##     I(NUMBRANCH^-6) + PCT_BLACK + FEMALE + COSTT4_A, data = train_set1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60187 -0.11933  0.01162  0.12620  0.38393
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.086e+00  4.335e-02  25.048  < 2e-16 ***
## AVGFACSAL       -3.190e-05  2.773e-06 -11.505  < 2e-16 ***
## CONTROL         -4.568e-02  1.514e-02  -3.016 0.002618 **
## POVERTY_RATE    -5.173e-03  8.832e-04  -5.857 6.19e-09 ***
## I(NUMBRANCH^-6) -7.197e-02  1.639e-02  -4.392 1.23e-05 ***
## PCT_BLACK       -2.639e-03  4.827e-04  -5.467 5.64e-08 ***
## FEMALE           2.338e-01  4.261e-02   5.486 5.08e-08 ***
## COSTT4_A        -2.089e-06  5.544e-07  -3.768 0.000173 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1672 on 1109 degrees of freedom
## Multiple R-squared:  0.2634, Adjusted R-squared:  0.2587
## F-statistic: 56.65 on 7 and 1109 DF,  p-value: < 2.2e-16
```