

project final copy

Amelie Devine

2023-12-07

```
library(olsrr)
```

```
##
```

```
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:datasets':
```

```
##
```

```
## rivers
```

```
library(readxl)
```

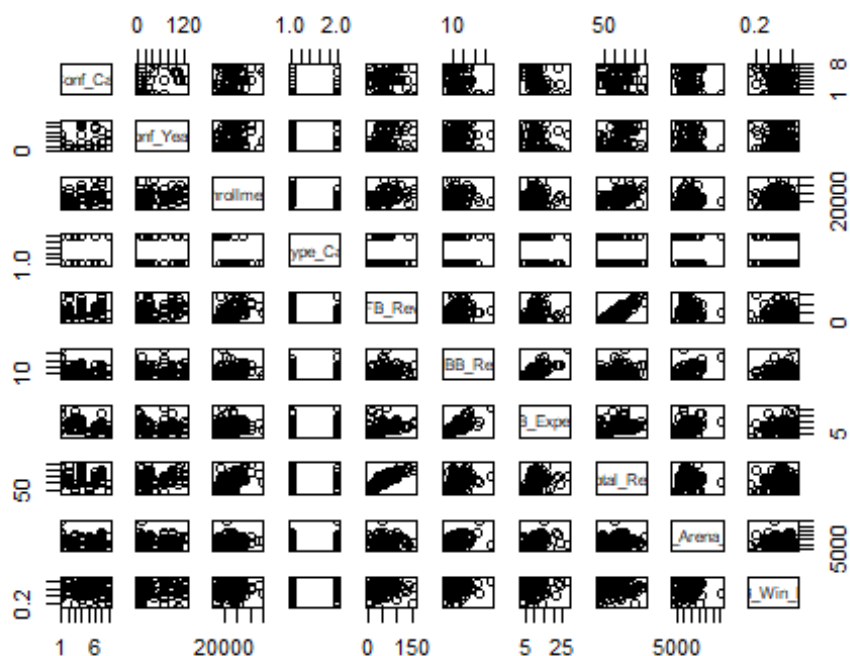
```
data_win = read_excel("proj_data_win.xlsx", na="-99")
```

```
data=data_win
```

```
summary(data)
```

```
##      Conf_Cat      Conf_Years      Enrollment      Type_Cat
##  Min.   :1.000    Min.   : 2.00    Min.   : 3200    Min.   :1.000
## 1st Qu.:2.000    1st Qu.: 10.00    1st Qu.:17593    1st Qu.:1.000
## Median :4.000    Median : 31.00    Median :27950    Median :1.000
## Mean   :4.417    Mean   : 48.98    Mean   :28946    Mean   :1.323
## 3rd Qu.:6.250    3rd Qu.: 90.00    3rd Qu.:36947    3rd Qu.:2.000
## Max.   :8.000    Max.   :126.00    Max.   :79232    Max.   :2.000
##
##      FB_Rev      MBB_Rev      MBB_Expense      Total_Rev
##  Min.   : 1.235    Min.   : 2.690    Min.   : 2.690    Min.   : 20.78
## 1st Qu.: 35.850    1st Qu.: 8.458    1st Qu.: 7.594    1st Qu.: 48.84
## Median : 52.242    Median :12.227    Median : 9.992    Median :108.46
## Mean   : 58.358    Mean   :13.247    Mean   :10.343    Mean   :106.63
## 3rd Qu.: 78.697    3rd Qu.:16.262    3rd Qu.:12.416    3rd Qu.:144.20
## Max.   :161.533    Max.   :45.109    Max.   :28.020    Max.   :246.61
## NA's   :15
## MBB_Arena_Cap      MBB_Win_Perc
##  Min.   : 3104    Min.   :0.0968
## 1st Qu.: 9454    1st Qu.:0.4801
## Median :12494    Median :0.5942
## Mean   :12817    Mean   :0.5802
## 3rd Qu.:15390    3rd Qu.:0.7154
## Max.   :35446    Max.   :0.8919
##
```

```
pairs(data)
```



```
fit = lm(MBB_Win_Perc~., data=data)
summary(fit)

##
## Call:
## lm(formula = MBB_Win_Perc ~ ., data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44357 -0.09320  0.00254  0.13578  0.29816
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.407e-01  1.287e-01   3.424  0.00103 **
## Conf_Cat       4.518e-03  1.005e-02   0.450  0.65442
## Conf_Years     2.257e-05  5.537e-04   0.041  0.96759
## Enrollment     2.555e-09  1.738e-06   0.001  0.99883
## Type_Cat       1.596e-02  5.645e-02   0.283  0.77817
## FB_Rev        -1.022e-03  1.226e-03  -0.834  0.40727
## MBB_Rev         8.970e-03  5.133e-03   1.747  0.08488 .
## MBB_Expense    -3.520e-03  7.637e-03  -0.461  0.64628
## Total_Rev       7.857e-04  1.007e-03   0.780  0.43795
## MBB_Arena_Cap  -1.725e-06  4.732e-06  -0.365  0.71650
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1674 on 71 degrees of freedom
```

```

## (15 observations deleted due to missingness)
## Multiple R-squared:  0.1151, Adjusted R-squared:  0.002878
## F-statistic: 1.026 on 9 and 71 DF,  p-value: 0.4284

anova(fit)

## Analysis of Variance Table
##
## Response: MBB_Win_Perc
##           Df Sum Sq Mean Sq F value Pr(>F)
## Conf_Cat    1 0.00390  0.003905   0.1394 0.70997
## Conf_Years   1 0.03670  0.036699   1.3104 0.25618
## Enrollment   1 0.00577  0.005772   0.2061 0.65123
## Type_Cat     1 0.01626  0.016259   0.5806 0.44862
## FB_Rev       1 0.00197  0.001975   0.0705 0.79138
## MBB_Rev       1 0.17394  0.173940   6.2107 0.01503 *
## MBB_Expense   1 0.00023  0.000231   0.0083 0.92782
## Total_Rev     1 0.01602  0.016022   0.5721 0.45193
## MBB_Arena_Cap 1 0.00372  0.003723   0.1329 0.71650
## Residuals    71 1.98848  0.028007
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

fit_best = ols_step_best_subset(lm(MBB_Win_Perc~., data=data))
final = data.frame(x=fit_best$predictors, r2=fit_best$rsquare, adjr2 =
fit_best$adjr,
                    AIC=fit_best$aic, BIC=fit_best$sbic,
r2press=fit_best$predrsq, Cp=fit_best$cp)
final

##
x
## 1
MBB_Rev
## 2
Type_Cat MBB_Rev
## 3
Conf_Cat Type_Cat MBB_Rev
## 4
Conf_Years Type_Cat MBB_Rev
## 5
MBB_Rev MBB_Expense
## 6
Total_Rev MBB_Arena_Cap
## 7
Conf_Cat Type_Cat FB_Rev MBB_Rev MBB_Expense
Total_Rev MBB_Arena_Cap
## 8
Conf_Cat Conf_Years Type_Cat FB_Rev MBB_Rev MBB_Expense
Total_Rev MBB_Arena_Cap
## 9 Conf_Cat Conf_Years Enrollment Type_Cat FB_Rev MBB_Rev MBB_Expense
Total_Rev MBB_Arena_Cap
##           r2      adjr2      AIC      BIC      r2press      Cp

```

```
## 1 0.1029232 0.093379792 -73.39222 -65.69918 0.06739601 -4.2247161
## 2 0.1091770 0.090019531 -72.06382 -61.80643 0.05015541 -2.8366298
## 3 0.1122669 0.083319121 -70.39739 -57.57564 0.03465980 -1.1389661
## 4 0.1130322 0.074044600 -68.48018 -53.09409 0.01385079 0.7861566
## 5 0.1133328 0.064073514 -66.51272 -48.56228 -0.01340984 2.7567434
## 6 0.1137947 0.041940162 -54.29030 -35.13471 -0.07828786 4.1010231
## 7 0.1150331 0.030173247 -52.40358 -30.85354 -0.12876150 6.0016622
## 8 0.1150538 0.016726421 -50.40547 -26.46098 -0.16017053 8.0000022
## 9 0.1150538 0.002877527 -48.40547 -22.06653 -0.18417435 10.0000000
```

Best R^2 : model 8 & 9 (highest) Best R^2_{adj} : model 1 (highest) Best R^2_{press} : model 1 (highest)
 Best C_p : model 8 (closest to $k+1$) Best AIC: model 1 (lowest) Best BIC: model 1 (lowest)

We choose model 1 as our best model because it fits 4/6 optimality criterion.

```
fit_step = ols_step_both_p(lm(MBB_Win_Perc~., data=data), prem=0.30,
pent=0.15, details=T)
```

```
## Stepwise Selection Method
## -----
##
## Candidate Terms:
##
## 1. Conf_Cat
## 2. Conf_Years
## 3. Enrollment
## 4. Type_Cat
## 5. FB_Rev
## 6. MBB_Rev
## 7. MBB_Expense
## 8. Total_Rev
## 9. MBB_Arena_Cap
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - MBB_Rev added
##
##                                     Model Summary
## -----
## R                                0.321          RMSE          0.162
## R-Squared                        0.103          Coef. Var      27.871
## Adj. R-Squared                   0.093          MSE           0.026
## Pred R-Squared                   0.067          MAE           0.129
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
```

```

##
##                                     ANOVA
## -----
##               Sum of
##               Squares           DF      Mean Square           F           Sig.
## -----
## Regression      0.282             1           0.282        10.785        0.0014
## Residual        2.458            94           0.026
## Total           2.740            95
## -----
##
##                                     Parameter Estimates
## -----
##
##               model      Beta      Std. Error      Std. Beta      t           Sig
## lower      upper
## -----
## (Intercept)    0.482          0.034                  14.177        0.000
## 0.415      0.550
## MBB_Rev        0.007          0.002           0.321         3.284        0.001
## 0.003      0.012
## -----
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## -----
##
##                                     Model Summary
## -----
## R                0.321          RMSE                0.162
## R-Squared         0.103          Coef. Var         27.871
## Adj. R-Squared    0.093          MSE                0.026
## Pred R-Squared    0.067          MAE                0.129
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                                     ANOVA
## -----
##               Sum of
##               Squares           DF      Mean Square           F           Sig.
## -----
## Regression      0.282             1           0.282        10.785        0.0014

```

```
## Residual      2.458      94      0.026
## Total        2.740      95
## -----
##
##                      Parameter Estimates
## -----
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig
lower      upper
## -----
## -----
## (Intercept)    0.482      0.034      14.177      0.000
0.415    0.550
##      MBB_Rev    0.007      0.002      0.321      3.284      0.001
0.003    0.012
## -----
## -----
```

Stepwise selection chose model 1

```
fit_back = ols_step_backward_p(lm(MBB_Win_Perc~., data=data), prem=0.15,
details=T)
```

```
## Backward Elimination Method
## -----
##
## Candidate Terms:
##
## 1 . Conf_Cat
## 2 . Conf_Years
## 3 . Enrollment
## 4 . Type_Cat
## 5 . FB_Rev
## 6 . MBB_Rev
## 7 . MBB_Expense
## 8 . Total_Rev
## 9 . MBB_Arena_Cap
##
## We are eliminating variables based on p value...
##
## - Enrollment
##
## Backward Elimination: Step 1
##
## Variable Enrollment Removed
##
##                      Model Summary
## -----
## R                      0.339      RMSE                      0.166
## R-Squared              0.115      Coef. Var              28.691
```

```

## Adj. R-Squared      0.017      MSE      0.028
## Pred R-Squared     -0.160      MAE      0.127
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                Sum of      DF      Mean Square      F      Sig.
##                Squares
## -----
## Regression      0.259        8          0.032      1.17    0.3290
## Residual        1.988       72          0.028
## Total           2.247       80
## -----
##
##                               Parameter Estimates
## -----
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig
## lower      upper
## -----
## -----
##      (Intercept)      0.441          0.115              3.824      0.000
## 0.211      0.671
##      Conf_Cat      0.005          0.010          0.060      0.464      0.644      -
## 0.015      0.024
##      Conf_Years      0.000          0.001          0.006      0.041      0.967      -
## 0.001      0.001
##      Type_Cat      0.016          0.050          0.038      0.319      0.751      -
## 0.084      0.115
##      FB_Rev      -0.001          0.001          -0.219     -0.845      0.401      -
## 0.003      0.001
##      MBB_Rev      0.009          0.005          0.398      1.791      0.078      -
## 0.001      0.019
##      MBB_Expense     -0.004          0.007          -0.097     -0.473      0.637      -
## 0.018      0.011
##      Total_Rev      0.001          0.001          0.230      0.823      0.413      -
## 0.001      0.003
##      MBB_Arena_Cap     0.000          0.000          -0.048     -0.367      0.714
## 0.000      0.000
## -----
## -----
##
##
## - Conf_Years
##
## Backward Elimination: Step 2
##

```

Variable Conf_Years Removed

##

Model Summary

```
## -----
## R                0.339      RMSE                0.165
## R-Squared        0.115      Coef. Var            28.494
## Adj. R-Squared   0.030      MSE                 0.027
## Pred R-Squared   -0.129      MAE                 0.127
## -----
```

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

##

ANOVA

```
## -----
##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression    0.258        7          0.037      1.356    0.2372
## Residual      1.989       73          0.027
## Total         2.247       80
## -----
```

##

Parameter Estimates

```
## -----
##              model
##              lower  upper      Beta      Std. Error      Std. Beta      t      Sig
## -----
## (Intercept)    0.213  0.669    0.441      0.114              3.854    0.000
## Conf_Cat       0.013  0.023    0.005      0.009      0.062    0.520    0.605    -
## Type_Cat       0.083  0.114    0.016      0.050      0.038    0.320    0.750    -
## FB_Rev         0.003  0.001   -0.001      0.001     -0.220   -0.860    0.393    -
## MBB_Rev        0.000  0.018    0.009      0.005      0.401    1.935    0.057
## MBB_Expense    0.018  0.010   -0.004      0.007     -0.099   -0.511    0.611    -
## Total_Rev      0.001  0.003    0.001      0.001      0.233    0.868    0.388    -
## MBB_Arena_Cap  0.000  0.000    0.000      0.000     -0.049   -0.382    0.704
## -----
```

##

##


```

## - Type_Cat
##
## Backward Elimination: Step 3
##
## Variable Type_Cat Removed
##
##                               Model Summary
## -----
## R                               0.337          RMSE              0.164
## R-Squared                       0.114          Coef. Var        28.321
## Adj. R-Squared                   0.042          MSE              0.027
## Pred R-Squared                   -0.078          MAE              0.127
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      0.256          6          0.043      1.584      0.1639
## Residual        1.991          74          0.027
## Total           2.247          80
## -----
##
##                               Parameter Estimates
## -----
## -----
## model      Beta      Std. Error      Std. Beta      t      Sig
lower  upper
## -----
## (Intercept)  0.465          0.084          5.520      0.000
0.297  0.633
## Conf_Cat     0.004          0.009          0.055      0.471      0.639      -
0.013  0.022
## FB_Rev       -0.001          0.001          -0.230     -0.914      0.364      -
0.003  0.001
## MBB_Rev       0.009          0.005          0.403      1.956      0.054
0.000  0.018
## MBB_Expense  -0.003          0.007          -0.094     -0.492      0.624      -
0.017  0.011
## Total_Rev     0.001          0.001          0.236      0.886      0.378      -
0.001  0.003
## MBB_Arena_Cap 0.000          0.000          -0.058     -0.463      0.645
0.000  0.000
## -----
## -----

```

```

##
##
## - MBB_Arena_Cap
##
## Backward Elimination: Step 4
##
## Variable MBB_Arena_Cap Removed
##
##
## Model Summary
## -----
## R                0.334      RMSE                0.163
## R-Squared        0.111      Coef. Var          28.172
## Adj. R-Squared   0.052      MSE                0.027
## Pred R-Squared   -0.044      MAE                0.127
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
## ANOVA
## -----
## Sum of Squares      DF      Mean Square      F      Sig.
## -----
## Regression          0.250         5          0.050      1.877      0.1084
## Residual            1.997        75          0.027
## Total                2.247        80
## -----
##
## Parameter Estimates
## -----
## model      Beta      Std. Error      Std. Beta      t      Sig
## lower      upper
## -----
## (Intercept)      0.445         0.071             6.279      0.000
## 0.303      0.586
## Conf_Cat         0.004         0.009         0.049      0.428      0.670      -
## 0.014      0.021
## FB_Rev          -0.001         0.001        -0.220     -0.880      0.381      -
## 0.003      0.001
## MBB_Rev          0.008         0.004         0.363      1.954      0.054
## 0.000      0.017
## MBB_Expense      -0.003         0.007        -0.080     -0.423      0.673      -
## 0.017      0.011
## Total_Rev        0.001         0.001         0.234      0.882      0.380      -
## 0.001      0.003
## -----
## -----

```

```

##
##
## - MBB_Expense
##
## Backward Elimination: Step 5
##
## Variable MBB_Expense Removed
##
##
## Model Summary
## -----
## R                0.330      RMSE                0.162
## R-Squared        0.109      Coef. Var          28.020
## Adj. R-Squared   0.062      MSE                0.026
## Pred R-Squared   -0.012     MAE                0.127
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
## ANOVA
## -----
## Sum of
## Squares      DF      Mean Square      F      Sig.
## -----
## Regression    0.245      4      0.061      2.327      0.0638
## Residual      2.002     76      0.026
## Total         2.247     80
## -----
##
## Parameter Estimates
## -----
##
## model      Beta      Std. Error      Std. Beta      t      Sig
## lower      upper
## -----
## (Intercept) 0.438      0.069      6.372      0.000
## 0.301      0.575
## Conf_Cat    0.004      0.009      0.049      0.434      0.665      -
## 0.013      0.021
## FB_Rev     -0.001      0.001     -0.186     -0.791      0.432      -
## 0.003      0.001
## MBB_Rev     0.007      0.003      0.306      2.402      0.019
## 0.001      0.013
## Total_Rev   0.001      0.001      0.195      0.788      0.433      -
## 0.001      0.002
## -----
##
##
##

```

```

## - Conf_Cat
##
## Backward Elimination: Step 6
##
## Variable Conf_Cat Removed
##
##                               Model Summary
## -----
## R                               0.327          RMSE              0.161
## R-Squared                       0.107          Coef. Var        27.872
## Adj. R-Squared                   0.072          MSE              0.026
## Pred R-Squared                   0.007          MAE              0.127
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      0.240           3           0.080      3.072      0.0326
## Residual        2.007          77           0.026
## Total           2.247          80
## -----
##
##                               Parameter Estimates
## -----
## -----
## model      Beta      Std. Error      Std. Beta      t      Sig
lower      upper
## -----
## (Intercept)  0.456           0.054              8.415      0.000
0.348      0.564
## FB_Rev      -0.001           0.001          -0.183     -0.781      0.437      -
0.003      0.001
## MBB_Rev      0.007           0.003           0.289      2.396      0.019
0.001      0.012
## Total_Rev    0.001           0.001           0.199      0.812      0.419      -
0.001      0.002
## -----
## -----
##
##
## - FB_Rev
##
## Backward Elimination: Step 7
##

```

Variable FB_Rev Removed

##

Model Summary

```
## -----
## R                0.322      RMSE                0.162
## R-Squared        0.104      Coef. Var            28.005
## Adj. R-Squared   0.085      MSE                 0.026
## Pred R-Squared   0.047      MAE                 0.128
## -----
```

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

##

ANOVA

```
## -----
##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression    0.285        2          0.142      5.392    0.0061
## Residual      2.456       93          0.026
## Total         2.740       95
## -----
```

##

Parameter Estimates

```
## -----
##              model      Beta      Std. Error      Std. Beta      t      Sig
## lower      upper
## -----
## (Intercept)  0.489      0.040              12.276    0.000
## 0.410      0.568
## MBB_Rev      0.008      0.003          0.337      3.038    0.003
## 0.003      0.013
## Total_Rev    0.000      0.000         -0.035     -0.319    0.750    -
## 0.001      0.001
## -----
```

##

##

##

- Total_Rev

##

Backward Elimination: Step 8

##

Variable Total_Rev Removed

##

Model Summary

```
## -----
## R                0.321      RMSE                0.162
## R-Squared        0.103      Coef. Var            27.871
```

```

## Adj. R-Squared      0.093      MSE      0.026
## Pred R-Squared     0.067      MAE      0.129
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression      0.282        1          0.282      10.785      0.0014
## Residual        2.458       94          0.026
## Total           2.740       95
## -----
##
##                               Parameter Estimates
## -----
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig
## lower      upper
## -----
## -----
## (Intercept)      0.482          0.034          14.177      0.000
## 0.415      0.550
## MBB_Rev      0.007          0.002          0.321      3.284      0.001
## 0.003      0.012
## -----
## -----
##
##
##
## No more variables satisfy the condition of p value = 0.15
##
##
## Variables Removed:
##
## - Enrollment
## - Conf_Years
## - Type_Cat
## - MBB_Arena_Cap
## - MBB_Expense
## - Conf_Cat
## - FB_Rev
## - Total_Rev
##
##
## Final Model Output
## -----

```

```
##
##                               Model Summary
## -----
## R                               0.321          RMSE                0.162
## R-Squared                       0.103          Coef. Var          27.871
## Adj. R-Squared                   0.093          MSE                 0.026
## Pred R-Squared                   0.067          MAE                 0.129
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      0.282          1          0.282      10.785      0.0014
## Residual        2.458          94          0.026
## Total           2.740          95
## -----
##
##                               Parameter Estimates
## -----
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig
## lower      upper
## -----
## (Intercept)    0.482          0.034                14.177      0.000
## 0.415      0.550
## MBB_Rev         0.007          0.002          0.321      3.284      0.001
## 0.003      0.012
## -----
## -----
```

Backward Regression chose Model 1

#New Model

```
new_fit = lm(MBB_Win_Perc~MBB_Rev, data=data)
summary(new_fit)

##
## Call:
## lm(formula = MBB_Win_Perc ~ MBB_Rev, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44098 -0.10004  0.00514  0.13058  0.28528
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.482490   0.034033  14.177 < 2e-16 ***
## MBB_Rev      0.007378   0.002247   3.284 0.00144 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1617 on 94 degrees of freedom
## Multiple R-squared:  0.1029, Adjusted R-squared:  0.09338
## F-statistic: 10.78 on 1 and 94 DF,  p-value: 0.001438

anova(new_fit)

## Analysis of Variance Table
##
## Response: MBB_Win_Perc
##           Df Sum Sq Mean Sq F value    Pr(>F)
## MBB_Rev     1 0.28204 0.282045   10.785 0.001438 **
## Residuals  94 2.45830 0.026152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#Hypothesis Test on Full vs Reduced Model H_0 : Reduced model is adequate. H_a : Full model is adequate.

```
test_F = ((2.2458-1.988)/7)/(1.988/(96-9))
crit_F = qf(0.05,7,87,lower.tail=F)
```

F test statistic 1.6117 > F critical value 2.1167 so we fail to reject the null hypothesis. We do not have evidence to show the full model is adequate.

#Hypothesis Test for Significance of Estimates $H_0: B_j = 0$; the j^{th} predictor is not significant. $H_a: B_j \neq 0$; the j^{th} predictor is significant. Predictor MBB_Rev has a p-value of 0.00144, so we reject H_0 and find that it is a significant predictor of men's basketball winning percentage.

#Predict using NOVA!

```
test = data.frame(MBB_Rev=20.513366)
predict(new_fit, test)
```

```
##           1
## 0.6338475
```

Prediction was 63.38%, actual was 78.95% (19.7% error)

#Confidence Interval

```
predict(new_fit, test, interval='confidence')
```

```
##           fit          lwr          upr
## 1 0.6338475 0.5877517 0.6799433
```


#Prediction Interval

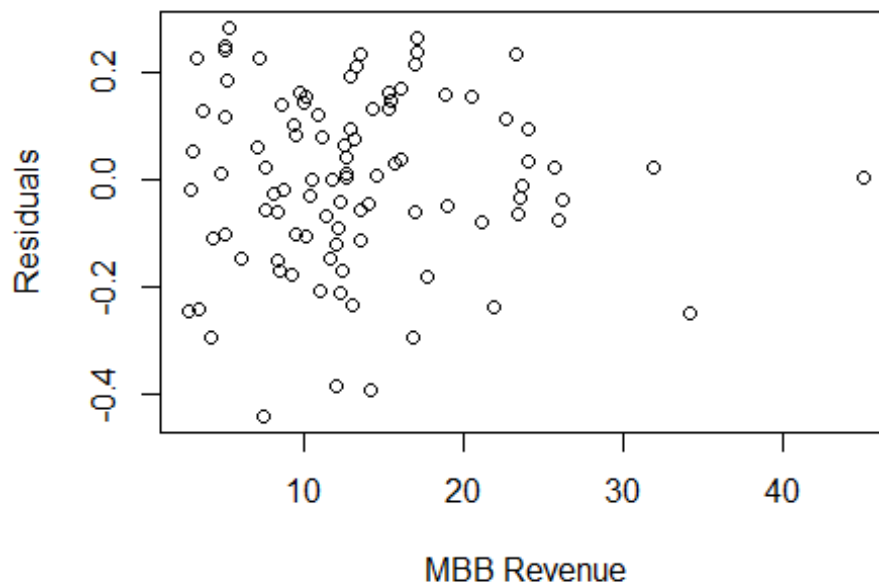
```
predict(new_fit, test, interval='prediction')
```

```
##           fit           lwr           upr  
## 1 0.6338475 0.3094646 0.9582305
```

CHECKING ASSUMPTIONS

#Linearity

```
ei = new_fit$residuals  
plot(data$MBB_Rev, ei, xlab='MBB Revenue', ylab='Residuals')
```

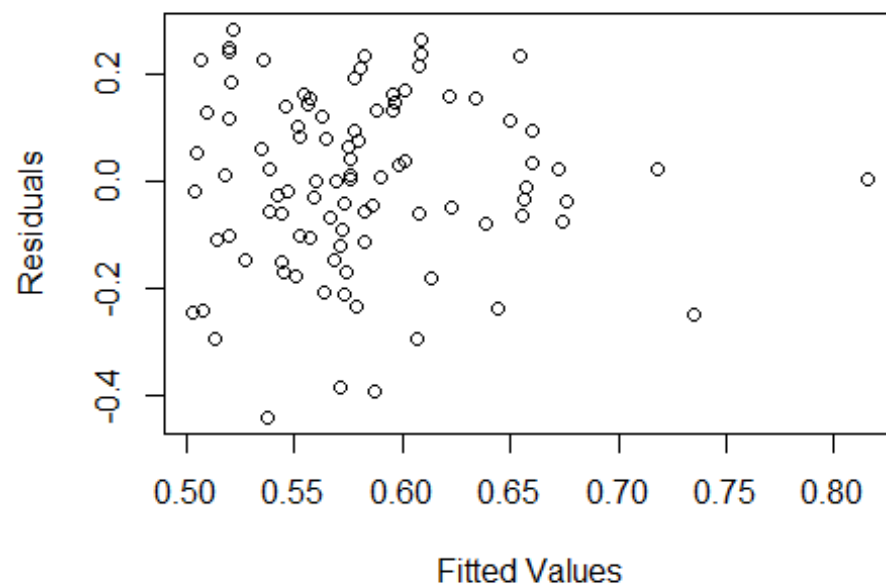


No obvious

systematic pattern

#Equal Variances

```
plot(new_fit$fitted.values, ei, xlab='Fitted Values', ylab='Residuals')
```

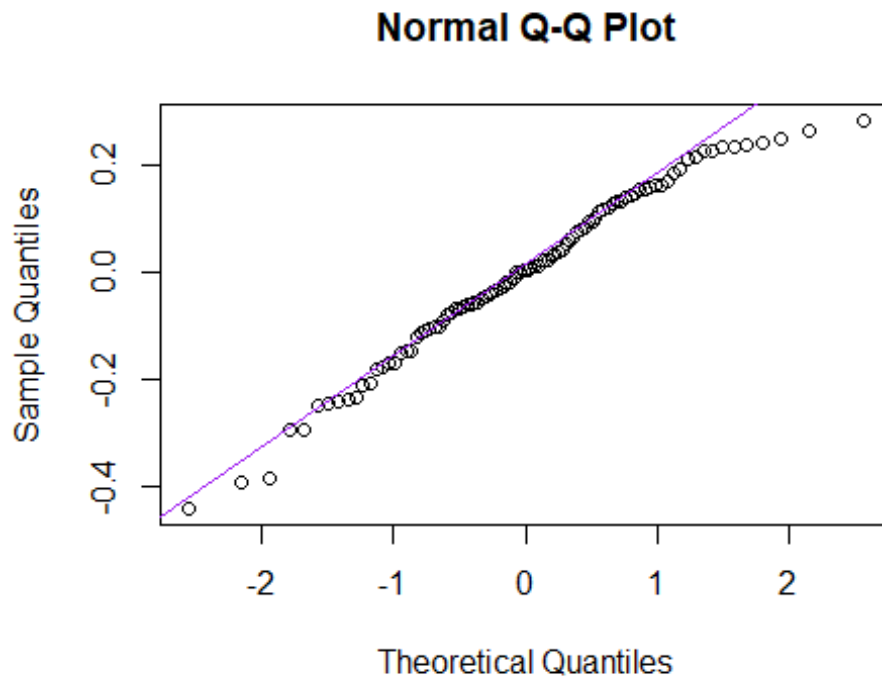


no systematic

pattern!

#Normality

```
qqnorm(ei)  
qqline(ei, col='purple')
```



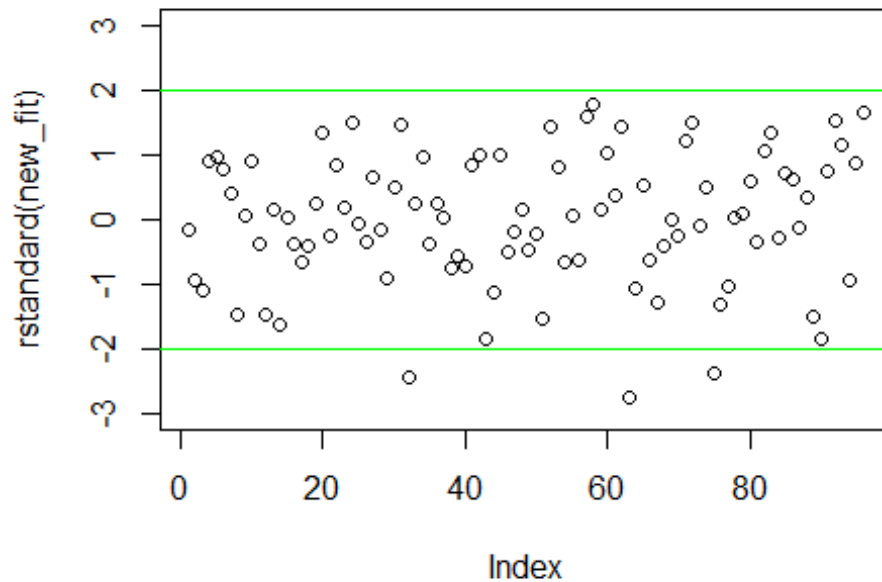
```
shapiro.test(ei)

##
##  Shapiro-Wilk normality test
##
## data:  ei
## W = 0.97759, p-value = 0.09907
```

normality satisfied: fit relatively close to line, but Wilk-Shapiro has high p-value

#Outliers

```
plot(rstandard(new_fit), ylim=c(-3,3))
abline(h=c(-2,2), col='green')
```



`rstandard(new_fit)`

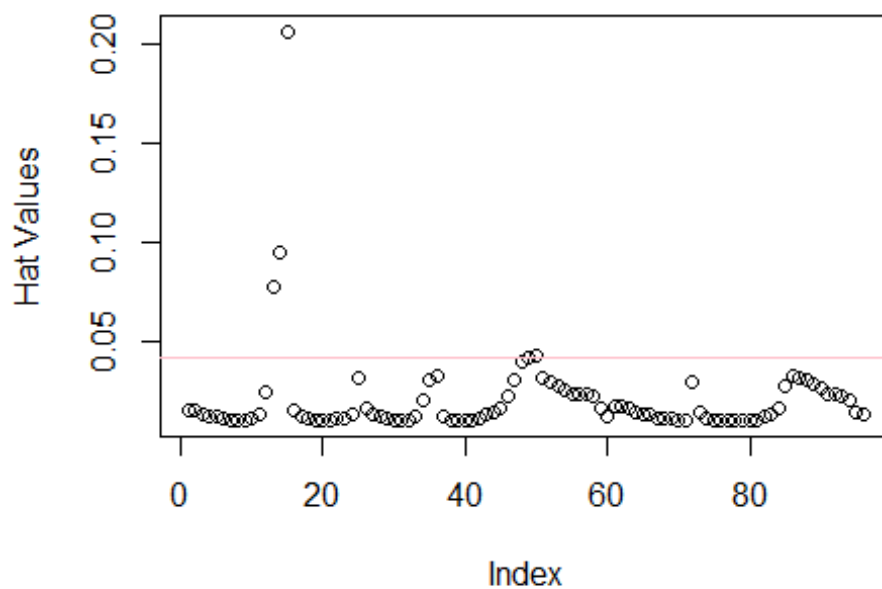
```
##          1          2          3          4          5          6
## -0.16565469 -0.93513415 -1.09452815  0.91214351  0.98045774  0.76747632
##          7          8          9         10         11         12
##  0.40095665 -1.45693715  0.06089641  0.91819383 -0.36993419 -1.48671249
##         13         14         15         16         17         18
##  0.16265360 -1.62304083  0.03593494 -0.37046013 -0.65879494 -0.41235976
##         19         20         21         22         23         24
##  0.25873186  1.32658953 -0.26820597  0.83570112  0.18934777  1.50115902
##         25         26         27         28         29         30
## -0.06154995 -0.34155576  0.65507122 -0.16996456 -0.89830574  0.48134842
##         31         32         33         34         35         36
##  1.46534058 -2.44623815  0.23548828  0.97257887 -0.38627398  0.23278821
##         37         38         39         40         41         42
##  0.01563358 -0.74298486 -0.55034628 -0.70629160  0.83422502  1.00974414
##         43         44         45         46         47         48
## -1.83137205 -1.11822235  1.00876432 -0.49834265 -0.19752419  0.15900894
##         49         50         51         52         53         54
## -0.46742218 -0.23661781 -1.53499415  1.43925444  0.81156231 -0.67492159
##         55         56         57         58         59         60
##  0.07058356 -0.62469835  1.57577439  1.78450386  0.14986120  1.02469449
##         61         62         63         64         65         66
##  0.37718601  1.42715241 -2.75007043 -1.05978022  0.52405838 -0.62672817
##         67         68         69         70         71         72
## -1.29594075 -0.41276877  0.01240990 -0.25654706  1.20465082  1.49159461
##         73         74         75         76         77         78
```

```
## -0.10564947  0.51091592 -2.38578690 -1.30082878 -1.04086527  0.03176607
##          79          80          81          82          83          84
##  0.07518434  0.60604093 -0.34154453  1.05970469  1.34154135 -0.29458850
##          85          86          87          88          89          90
##  0.72087388  0.60705113 -0.12270238  0.33829631 -1.51261605 -1.84596184
##          91          92          93          94          95          96
##  0.73236695  1.53135257  1.15892635 -0.92566428  0.87205561  1.65974477
```

3 outliers: 32, 63, 75

#Influential Points

```
bm = 2*(2/96) ## about 0.0417
plot(hatvalues(new_fit), ylab='Hat Values')
abline(h=bm, col='pink')
```



```
hatvalues(new_fit)
##          1          2          3          4          5          6
7
##  0.01565888  0.01507526  0.01348651  0.01247856  0.01237500  0.01155578
0.01053713
##          8          9         10         11         12         13
14
##  0.01043160  0.01076945  0.01129327  0.01311760  0.02471560  0.07804894
0.09481776
##         15         16         17         18         19         20
21
```

```
## 0.20637368 0.01500454 0.01225123 0.01110330 0.01048075 0.01041741
0.01053651
##          22          23          24          25          26          27
28
## 0.01122337 0.01155608 0.01332303 0.03125653 0.01653882 0.01342981
0.01207250
##          29          30          31          32          33          34
35
## 0.01088989 0.01041780 0.01043446 0.01057906 0.01195294 0.02060944
0.03040268
##          36          37          38          39          40          41
42
## 0.03296646 0.01186961 0.01071175 0.01063264 0.01043369 0.01063134
0.01121400
##          43          44          45          46          47          48
49
## 0.01290921 0.01418862 0.01661066 0.02244698 0.03101089 0.04015078
0.04159246
##          50          51          52          53          54          55
56
## 0.04316034 0.03192797 0.02992246 0.02814541 0.02595972 0.02409252
0.02361232
##          57          58          59          60          61          62
63
## 0.02345930 0.02276279 0.01660066 0.01282704 0.01794433 0.01737776
0.01680651
##          64          65          66          67          68          69
70
## 0.01478947 0.01319988 0.01318618 0.01144843 0.01109681 0.01083223
0.01062489
##          71          72          73          74          75          76
77
## 0.01044104 0.02987701 0.01448383 0.01124530 0.01070260 0.01060999
0.01056634
##          78          79          80          81          82          83
84
## 0.01049634 0.01047699 0.01044297 0.01043908 0.01195443 0.01314731
0.01691302
##          85          86          87          88          89          90
91
## 0.02754733 0.03313068 0.03132941 0.03051853 0.02905039 0.02626363
0.02355034
##          92          93          94          95          96
## 0.02329697 0.02300220 0.02028668 0.01464149 0.01322485
```

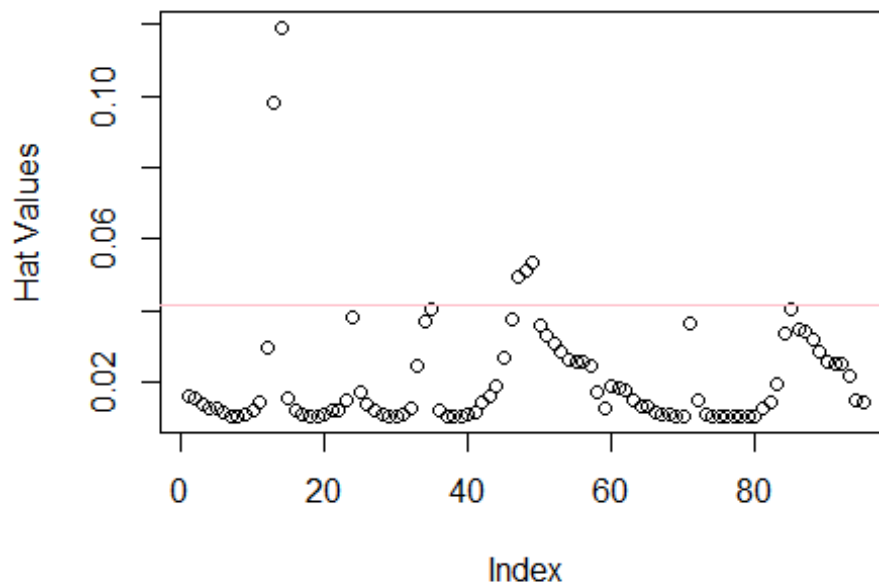
#Remove 15 - high influence point

```
data_2 <- data[-c(15),]
new_fit_2 = lm(MBB_Win_Perc~MBB_Rev, data=data_2)
summary(new_fit_2)
```

```
##
## Call:
## lm(formula = MBB_Win_Perc ~ MBB_Rev, data = data_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44114 -0.10043  0.00515  0.13168  0.28503
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.482953   0.036590  13.199 < 2e-16 ***
## MBB_Rev      0.007338   0.002522   2.909  0.00453 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1626 on 93 degrees of freedom
## Multiple R-squared:  0.08342,    Adjusted R-squared:  0.07357
## F-statistic: 8.464 on 1 and 93 DF,  p-value: 0.00453
```

#Recheck influential points

```
bm2 = 2*(2/95) ## about 0.0421
plot(hatvalues(new_fit_2), ylab='Hat Values')
abline(h=bm, col='pink')
```



```
hatvalues(new_fit_2)
```

##	1	2	3	4	5	6
7						
##	0.01624861	0.01556911	0.01373737	0.01259673	0.01248103	0.01158156
	0.01057605					
##	8	9	10	11	12	13
14						
##	0.01052711	0.01121155	0.01199049	0.01452514	0.02977247	0.09790702
	0.11917020					
##	15	16	17	18	19	20
21						
##	0.01548697	0.01234321	0.01110507	0.01054027	0.01056437	0.01083004
	0.01188933					
##	22	23	24	25	26	27
28						
##	0.01236639	0.01480383	0.03821642	0.01727795	0.01367264	0.01214521
	0.01089067					
##	29	30	31	32	33	34
35						
##	0.01054243	0.01062459	0.01090397	0.01292446	0.02443609	0.03711702
	0.04041603					
##	36	37	38	39	40	41
42						
##	0.01192214	0.01072172	0.01065192	0.01062256	0.01099133	0.01187572
	0.01424153					
##	43	44	45	46	47	48
49						
##	0.01597038	0.01919131	0.02682871	0.03790021	0.04963304	0.05147869
	0.05348466					
##	50	51	52	53	54	55
56						
##	0.03567178	0.03325248	0.03111234	0.02848549	0.02624705	0.02567236
	0.02548932					
##	57	58	59	60	61	62
63						
##	0.02465675	0.01735048	0.01298840	0.01893152	0.01826374	0.01759208
	0.01523747					
##	64	65	66	67	68	69
70						
##	0.01341077	0.01339519	0.01146665	0.01109841	0.01083467	0.01064534
	0.01052641					
##	71	72	73	74	75	76
77						
##	0.03643981	0.01488370	0.01125212	0.01071342	0.01063288	0.01059785
	0.01054901					
##	78	79	80	81	82	83
84						
##	0.01053835	0.01052659	0.01063634	0.01292654	0.01456550	0.01959038
	0.03343477					
##	85	86	87	88	89	90
91						

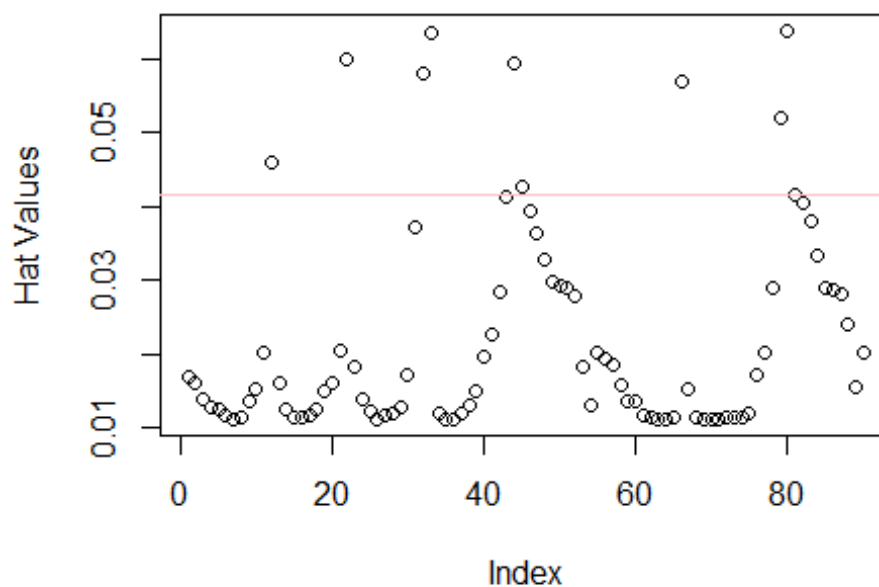

```
## 0.04062714 0.03494931 0.03397111 0.03220177 0.02885034 0.02559822
0.02529519
##          92          93          94          95
## 0.02494282 0.02170598 0.01506607 0.01467075
```

high influence points: 13, 14, 47, 48, 49

#Remove influence points and recheck again!

```
data_3 <- data_2[-c(13,14,47,48,49),]
new_fit_3 = lm(MBB_Win_Perc~MBB_Rev, data=data_3)

bm3 = 2*(2/90) ## about 0.0444
plot(hatvalues(new_fit_3), ylab='Hat Values')
abline(h=bm, col='pink')
```



#Compare first reduced model & model without 6 influential points

```
summary(new_fit)

##
## Call:
## lm(formula = MBB_Win_Perc ~ MBB_Rev, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44098 -0.10004  0.00514  0.13058  0.28528
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.482490   0.034033  14.177 < 2e-16 ***
## MBB_Rev      0.007378   0.002247   3.284 0.00144 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1617 on 94 degrees of freedom
## Multiple R-squared:  0.1029, Adjusted R-squared:  0.09338
## F-statistic: 10.78 on 1 and 94 DF,  p-value: 0.001438
```

```
summary(new_fit_2)
```

```
##
## Call:
## lm(formula = MBB_Win_Perc ~ MBB_Rev, data = data_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44114 -0.10043  0.00515  0.13168  0.28503
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.482953   0.036590  13.199 < 2e-16 ***
## MBB_Rev      0.007338   0.002522   2.909 0.00453 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1626 on 93 degrees of freedom
## Multiple R-squared:  0.08342,    Adjusted R-squared:  0.07357
## F-statistic: 8.464 on 1 and 93 DF,  p-value: 0.00453
```

```
summary(new_fit_3)
```

```
##
## Call:
## lm(formula = MBB_Win_Perc ~ MBB_Rev, data = data_3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.43439 -0.09733  0.00161  0.13247  0.29680
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.459415   0.041400  11.097 <2e-16 ***
## MBB_Rev      0.009579   0.003127   3.064 0.0029 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1642 on 88 degrees of freedom
```

```
## Multiple R-squared:  0.09639,    Adjusted R-squared:  0.08612  
## F-statistic: 9.387 on 1 and 88 DF,  p-value: 0.0029
```

fit1 and fit2 very similar, fit3 differ slightly, but not a ton and both (slope & intercept) still significant