Final Total Points Models

Import Libraries

```
library(readr)
library(lmerTest)
```

```
## Loading required package: lme4

## Loading required package: Matrix

##
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':
##
##     lmer

## The following object is masked from 'package:stats':
##
##     step
```

```
library(bestglm)
```

```
## Loading required package: leaps
```

```
library(leaps)
library(glmnet)
```

```
## Loaded glmnet 4.1-7
```

```
library(MASS)
```

Import Data

```
days = read_csv('../Datasets/For_Henoc.csv')$Day
```

```
## New names:
## Rows: 1070 Columns: 16
## -- Column specification
## -------------------------------------------------------- Delimiter: "," dbl
## (16): ...1, Home ID, Away ID, Day, Home_pts, Away_pts, Spread_Pred, Spre...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
games = read_csv('../Datasets/Final_Data.csv')[-1]
```

```
## New names:
## Rows: 1070 Columns: 52
## -- Column specification
## -------------------------------------------------------- Delimiter: "," dbl
## (52): ...1, Home_ID, Away_ID, Day, Home_pts, Away_pts, Spread_Pred.x, Sp...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
games = games[-6]
colnames(games)[6] = "Spread_Pred"
games$Day = days

games = na.omit(games)

# Specify Factor Data type
factor_vars = c("Home_ID","Away_ID","Left_ID","Right_ID","Matchup_ID")
games[factor_vars] = lapply(games[factor_vars],factor)

games$Total_Points = games$Home_pts + games$Away_pts

KnownHomeAwayStats = games[c(51,1,2,3:6,16:28)]
KnownLeftRightStats = games[c(51,3,28,31,32,39:50)]
```

Define MAE Function

```
MAE = function (predicted, actual){
  abs_err = abs(actual-predicted)

  mae = mean(abs_err)
  return(mae)
}
```

Linear Models and Stepwise Selection

Home and Away Individual

Predict Home Points

```
HomePoints.LM1 = lm(Home_pts ~ . + as.factor(Home_ID) ,
                    ## Remove Total points and Away pts
                    data = KnownHomeAwayStats[-c(1,6)])

MAE(predict(HomePoints.LM1), KnownHomeAwayStats$Home_pts)
```

```
## [1] 6.084372
```

Pred Home points Includes everything except Away Points

MAE = 6.08

Predict Away Points

```
AwayPoints.LM1 = lm(Away_pts ~ .,
                    ## Remove Total points and Away pts
                    data = KnownHomeAwayStats[-c(1,5)])

MAE(predict(AwayPoints.LM1), KnownHomeAwayStats$Away_pts)
```

```
## [1] 5.828596
```

Pred away points Includes everything except home Points

MAE = 5.829

Add Models Together

```
MAE(predict(AwayPoints.LM1)+predict(HomePoints.LM1),
    KnownHomeAwayStats$Total_Points)
```

```
## [1] 9.928312
```

MAE for summing together is 9.93

Drop 1 Test

Drops one variable in model and gives F-Test P-value for comparing full model to model with 1 less (Significant p implies adding the variable accounts for variance not explained by all other variables)

```
drop1(HomePoints.LM1, test = "F")[c(5,6)]
```

Home Points

```
##                   F value   Pr(>F)
## <none>
## Home_ID
## Away_ID
## Day                0.8855 0.347116
## Spread_Pred        7.4297 0.006626 **
## Home_TPG           2.7249 0.099381 .
## Away_TPG           0.7447 0.388550
```

```
## Home_TSG            0.2107 0.646426
## Away_TSG            0.0358 0.850056
## Home_TOG            0.0455 0.831102
## Away_TOG            0.1686 0.681524
## Home_PPG            0.5928 0.441677
## Away_PPG            0.3415 0.559198
## Home_SPG            2.6643 0.103211
## Away_SPG            0.0133 0.908171
## Home_OPG            2.3941 0.122390
## Away_OPG            0.6721 0.412681
## Matchup_ID          0.9601 0.666684
## as.factor(Home_ID)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Home and away Team account for significant variance, Home_SPG also could account for some

```
drop1(AwayPoints.LM1, test = "F")[c(5,6)]
```

Away Points

```
##             F value    Pr(>F)
## <none>
## Home_ID
## Away_ID
## Day          1.9553 0.1625922
## Spread_Pred  0.0806 0.7766632
## Home_TPG     0.2676 0.6051371
## Away_TPG     0.2713 0.6027062
## Home_TSG     0.5344 0.4650706
## Away_TSG     0.0719 0.7886356
## Home_TOG     0.3826 0.5364777
## Away_TOG     4.4072 0.0362561 *
## Home_PPG     0.1310 0.7174963
## Away_PPG    11.2389 0.0008578 ***
## Home_SPG     1.9088 0.1676720
## Away_SPG     1.7839 0.1822459
## Home_OPG     5.5032 0.0193468 *
## Away_OPG     0.2808 0.5964129
## Matchup_ID   1.0868 0.1847439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Home and away explain significant variance in Away points, as well as predicted spread and Away_OPG, and potentially Away_TSG.

High quantity of low F-Values could indicate some variables are insignificant and account for little variability. Stepwise selection to minimize AIC is benificial.

Stepwise Selection (Minimize AIC)

```
null = lm(Home_pts ~ 1, data = KnownHomeAwayStats[-c(1,6)])

HomeStepmod = stepAIC(HomePoints.LM1, null,
        direction = "both", steps = 100,
        trace = 0)

drop1(HomeStepmod, test = "F")[c(5,6)]
```

Home Stepwise

```
##             F value    Pr(>F)
## <none>
## Home_ID      2.7171 3.362e-06 ***
## Spread_Pred 53.6188 5.098e-13 ***
## Home_TPG     4.1255  0.042514 *
## Away_TPG     4.6984  0.030432 *
## Home_PPG     3.1166  0.077811 .
## Away_PPG     9.1235  0.002589 **
## Home_SPG     2.9894  0.084126 .
## Away_SPG     1.9532  0.162559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
MAE(predict(HomeStepmod),KnownHomeAwayStats[-c(1,6)]$Home_pts)
```

```
## [1] 8.630519
```

MAE = 8.46212

```
null = lm(Away_pts ~ 1, data = KnownHomeAwayStats[-c(1,5)])

AwayStepmod = stepAIC(AwayPoints.LM1, null,
        direction = "both", steps = 100,
        trace = 0)

drop1(AwayStepmod, test = "F")[c(5,6)]
```

Away Stepwise

```
##             F value    Pr(>F)
## <none>
## Home_ID      3.2302 2.872e-08 ***
## Spread_Pred  5.0468   0.02489 *
## Away_TPG    18.7009 1.686e-05 ***
## Home_TSG     4.1433   0.04207 *
## Away_TSG     6.1395   0.01339 *
## Away_TOG     3.6714   0.05565 .
```

```
## Away_OPG     2.8302   0.09283 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
MAE(predict(AwayStepmod),KnownHomeAwayStats[-c(1,5)]$Away_pts)
```

```
## [1] 8.262423
```

MAE = 8.317 (down from 10.909)

```
MAE(predict(AwayStepmod) + predict(HomeStepmod), KnownHomeAwayStats$Total_Points)
```

MAE From Combined

```
## [1] 14.07492
```

MAE = 13.793 (down from 18.596)

Refit Model using all variables from sub-models

```
TotalPoints.LM0 = lm(Total_Points ~ Home_ID+Day+Spread_Pred
                     +Away_TSG+Away_PPG + Home_TSG,
                    ## Remove home points and Away pts
                    data = KnownHomeAwayStats[-c(5,6)])

MAE(predict(TotalPoints.LM0), KnownHomeAwayStats$Total_Points)
```

```
## [1] 14.15265
```

MAE = 14.28695

Predict Total Points Without Home/Away Individual Points

```
TotalPoints.LM1 = lm(Total_Points ~ .,
                    ## Remove home points and Away pts
                    data = KnownHomeAwayStats[-c(5,6)])

MAE(predict(TotalPoints.LM1), KnownHomeAwayStats$Total_Points)
```

```
## [1] 9.928312
```

MAE = 9.666349 (Same as initial combined MAE)

```
drop1(TotalPoints.LM1, test = "F")[c(5,6)]
```

Drop 1 Test

```
##             F value  Pr(>F)
## <none>
## Home_ID
## Away_ID
## Day          1.9470 0.16349
## Spread_Pred  3.4187 0.06501 .
## Home_TPG     1.7447 0.18711
## Away_TPG     0.6993 0.40338
## Home_TSG     0.0202 0.88708
## Away_TSG     0.0745 0.78497
## Home_TOG     0.2420 0.62298
## Away_TOG     2.1780 0.14059
## Home_PPG     0.4711 0.49279
## Away_PPG     5.3485 0.02112 *
## Home_SPG     3.2856 0.07045 .
## Away_SPG     0.4988 0.48033
## Home_OPG     5.3893 0.02064 *
## Away_OPG     0.6641 0.41549
## Matchup_ID   1.0209 0.41045
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Home ID and away ID are both contribute to significant variance not explained by other variables. Away OPG could also explain significant variance.

```
null = lm(Total_Points ~ 1, data = KnownHomeAwayStats[-c(5,6)])

TotalPointsStepmod = stepAIC(TotalPoints.LM1, null,
        direction = "both", steps = 100,
        trace = 0)

drop1(TotalPointsStepmod, test = "F")[c(5,6)]
```

Stepwise

```
##             F value    Pr(>F)
## <none>
## Home_ID      3.8317 8.152e-11 ***
## Spread_Pred 10.6857  0.001117 **
## Home_TPG     5.1619  0.023304 *
## Away_TPG     6.0053  0.014437 *
## Away_PPG     7.8495  0.005184 **
## Home_SPG     3.1174  0.077773 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
MAE(predict(TotalPointsStepmod),KnownHomeAwayStats[-c(5,6)]$Total_Points)
```

```
## [1] 14.07619
```

MAE is 13.70024 (down from 18.5955 but higher than best MAE from individual models with MAE of 13.793)

Left and Right

```
LR_Data = KnownLeftRightStats[-c(4,5)]
```

Left and right removes the need to analyze left or right individually, because there is no "home field advantage" that is accounted for. These models essentiually test the assumption that whoever the home or away team is, will have very little impact on the total points scored in the game.

Predict Total Points Without Left/Right Individual Points

```
TotalPoints.LM1 = lm(Total_Points ~ .,
                     ## Remove Left points and Right pts
                     data = KnownLeftRightStats[-c(4,5)])

MAE(predict(TotalPoints.LM1), LR_Data$Total_Points)
```

```
## [1] 10.20835
```

MAE = 10.979 (Already lwoer than best MAE with Home and Away)

```
drop1(TotalPoints.LM1, test = "F")[c(5,6)]
```

Drop 1 Test

```
##            F value   Pr(>F)
## <none>
## Day         1.0344 0.309563
## Matchup_ID  1.2493 0.006687 **
## Left_TPG    0.7756 0.378854
## Right_TPG   0.4673 0.494515
## Left_TSG    0.2807 0.596450
## Right_TSG   1.4025 0.236808
## Left_TOG    0.0081 0.928104
## Right_TOG   2.6003 0.107402
## Left_PPG    0.1821 0.669699
## Right_PPG   9.8051 0.001830 **
## Left_SPG    4.4645 0.035045 *
```

```
## Right_SPG   0.0011 0.974055
## Left_OPG    0.0960 0.756786
## Right_OPG   9.9198 0.001722 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

None of the variables explain significant variance that is not explained by other variables in the data set.

```
null = lm(Total_Points ~ 1, data = KnownLeftRightStats[-c(4,5)])

TotalPointsStepmod = stepAIC(TotalPoints.LM1, null,
        direction = "both", steps = 100,
        trace = 0)

drop1(TotalPointsStepmod, test = "F")[c(5,6)]
```

Stepwise

```
##           F value    Pr(>F)
## <none>
## Right_TPG 22.2256 2.767e-06 ***
## Left_TSG   3.3695 0.0667080 .
## Right_TSG  5.8517 0.0157393 *
## Right_TOG  6.3573 0.0118436 *
## Left_PPG  11.5416 0.0007073 ***
## Right_OPG  8.6825 0.0032872 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
MAE(predict(TotalPointsStepmod),KnownLeftRightStats[-c(4,5)]$Total_Points)
```

```
## [1] 14.84167
```

MAE is 15.024 (up from 12.645)

Random Effects Analysis

Day Random Effect

Left Right

```
TotalPoints.LR.LMER1 = lmer(Total_Points~.-Day + (1|as.factor(Day)),
                        data = LR_Data)

MAE(predict(TotalPoints.LR.LMER1), LR_Data$Total_Points)
```

```
## [1] 8.665664
```

```
VarCorr(TotalPoints.LR.LMER1)
```

```
##  Groups          Name        Std.Dev.
##  as.factor(Day) (Intercept) 12.858
##  Residual                   17.689
```

MAE = 9.6052

Total points changes roughly +- 7.8 day to day

Home Only

```
home_pred = KnownHomeAwayStats[-c(1,6)]
```

```
TotalPoints.H.LMER1 = lmer(Home_pts~.-Day + (1|as.factor(Day)),
                           data = home_pred)
```

```
## fixed-effect model matrix is rank deficient so dropping 29 columns / coefficients
```

```
MAE(predict(TotalPoints.H.LMER1), home_pred$Home_pts)
```

```
## [1] 5.34621
```

```
VarCorr(TotalPoints.H.LMER1)
```

```
##  Groups          Name        Std.Dev.
##  as.factor(Day) (Intercept)  0.81889
##  Residual                   11.11097
```

MAE = 5.16

Home points changes roughly +- 11.52 day to day

With previous best Home Points Model

```
TotalPoints.H.LMER2 = lmer(Home_pts~Home_ID+Away_ID+Spread_Pred+Home_TSG
                             + Day-Day + (1|as.factor(Day)),
                           data = home_pred)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
MAE(predict(TotalPoints.H.LMER2), home_pred$Home_pts)
```

```
## [1] 8.45457
```

```
VarCorr(TotalPoints.H.LMER2)
```

```
##  Groups          Name        Std.Dev.
##  as.factor(Day) (Intercept)  0.000
##  Residual                    11.061
```

MAE = 8.46 (Less than stepwise linear model from before)

Total points has no significant change Day-to-Day for home team.

Away Only

```
away_pred = KnownHomeAwayStats[-c(1,5)]
```

```
TotalPoints.A.LMER1 = lmer(Away_pts~.-Day + (1|as.factor(Day)),
                           data = away_pred)
```

```
## fixed-effect model matrix is rank deficient so dropping 29 columns / coefficients
```

```
MAE(predict(TotalPoints.A.LMER1), away_pred$Away_pts)
```

```
## [1] 4.789
```

```
VarCorr(TotalPoints.A.LMER1)
```

```
##  Groups          Name        Std.Dev.
##  as.factor(Day) (Intercept)  6.9867
##  Residual                    10.2805
```

MAE = 5.06

Away points changes roughly +- 7.62 day to day

With previous best Away Points Model

```
TotalPoints.A.LMER2 = lmer(Away_pts~Home_ID+Day+Spread_Pred+Away_TSG
                             + Away_PPG -Day + (1|as.factor(Day)),
                           data = away_pred)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
MAE(predict(TotalPoints.A.LMER2), away_pred$Away_pts)
```

```
## [1] 8.282278
```

```
VarCorr(TotalPoints.A.LMER2)
```

```
##  Groups          Name        Std.Dev.
##  as.factor(Day) (Intercept)  0.000
##  Residual                    10.879
```

MAE = 8.3579

Combine Models

```
MAE(predict(TotalPoints.A.LMER2)+predict(TotalPoints.H.LMER2),LR_Data$Total_Points)
```

```
## [1] 13.75105
```

MAE is 13.780

Refit using optimal vars

```
TotalPoints.LMER3 = lmer(KnownHomeAwayStats$Total_Points~
                           Home_ID+Away_ID+Spread_Pred+Home_TSG
                           + Away_TSG + Away_PPG
                           + Day-Day + (1|as.factor(Day)),
                         data = home_pred)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
MAE(predict(TotalPoints.LMER3), KnownHomeAwayStats$Total_Points)
```

```
## [1] 13.66831
```

```
VarCorr(TotalPoints.A.LMER2)
```

```
##  Groups          Name        Std.Dev.
##  as.factor(Day) (Intercept)  0.000
##  Residual                    10.879
```

MAE after refitting is 13.685