Yiluo Wang

MA678

12/11/2022

FINAL REPORT

Abstract:

This report used the data called English Premier League Players Dataset, 2017/18 to research that club player's market value and if value have relationship with club size, FPL point and FPL value.

Introduction:

This data set Contains data of market values, page views, age, position and FPL stats of 461 players of the 17/18 English Premier League. There are 20 clubs in the league, the big club is six, people called these club "big six". Each team have at least 15 players. The market value as on transfermrkt.com on July 20th, 2017. The interesting thing is this data have market value and FPL value. The FPL is a game that casts you in the role of a Fantasy manager of Premier League players, the FPL value means the value of this player in the game. You can pick 15 players in the league score points for your team based on their performances for their clubs in PL matches. I am force on if the player's market value has relationship with FPL game and also see difference between big club and normal club.

Method:

The mean idea is using multilevel model to get what I find. In this dataset, I have three levels. Level 1 are the players; level 2 are clubs and level 3 are big club. I want to find if a player has high FPL value and point, he will have higher market value. For big club, there players have higher value than normal club. The figure 1.1 is all player's value group by club. We can see that some club have high point but can't get much information.

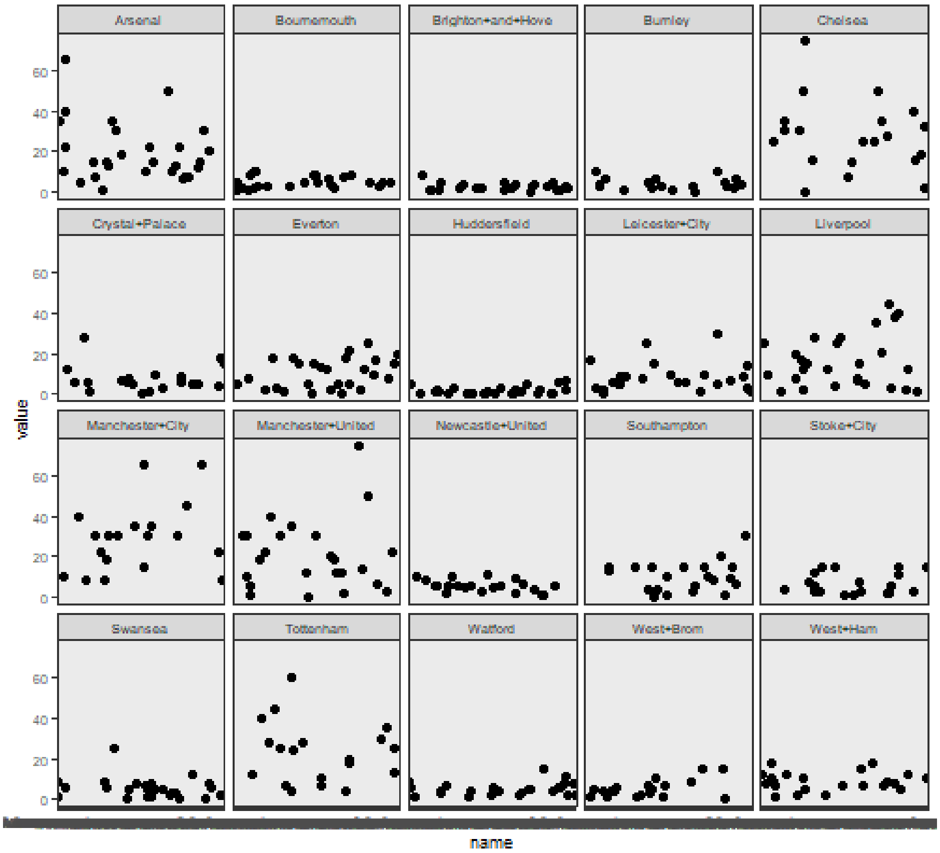

Figure 1.1

Then I want to Random intercept only Model, use model market_value~1+(1|club) to test and use icc to check if this explains any variance. Next, I use model market_value ~ fpl_points2 + fpl_value2 + (1|club) and market_value ~ fpl_points2*fpl_value2+(1|club) to test my fixed effects. Final I plot the random and fixed effects with fitted fpl_points. Figure 2 is the original one and figure 3 are the model one.

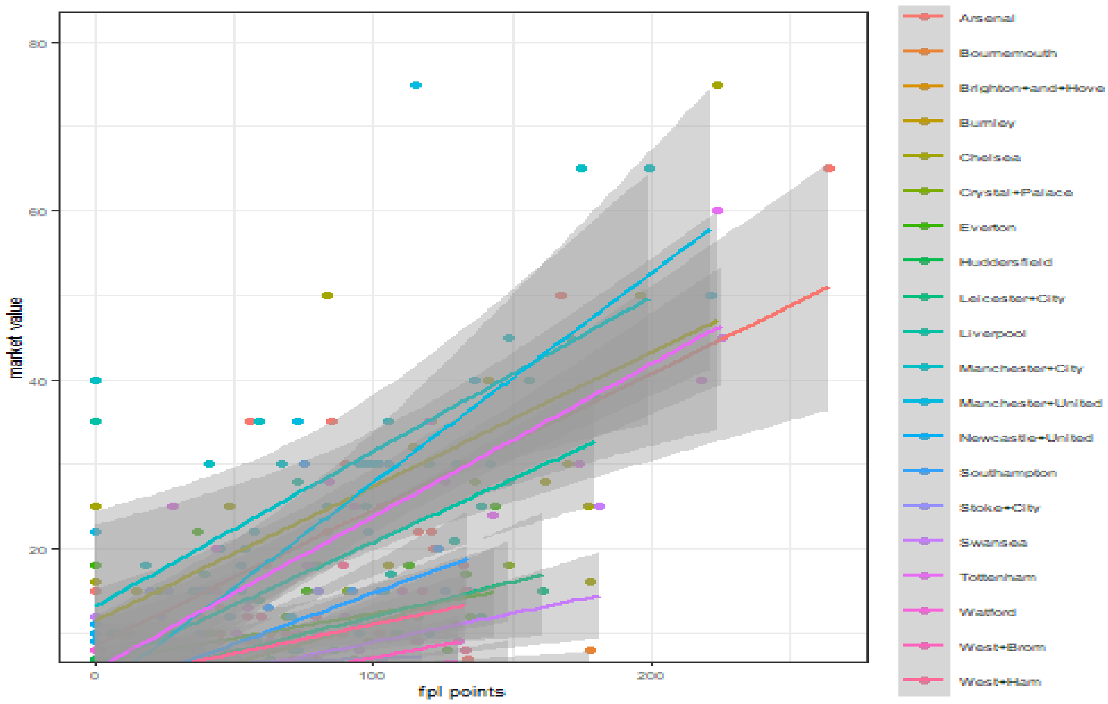

Figure 2

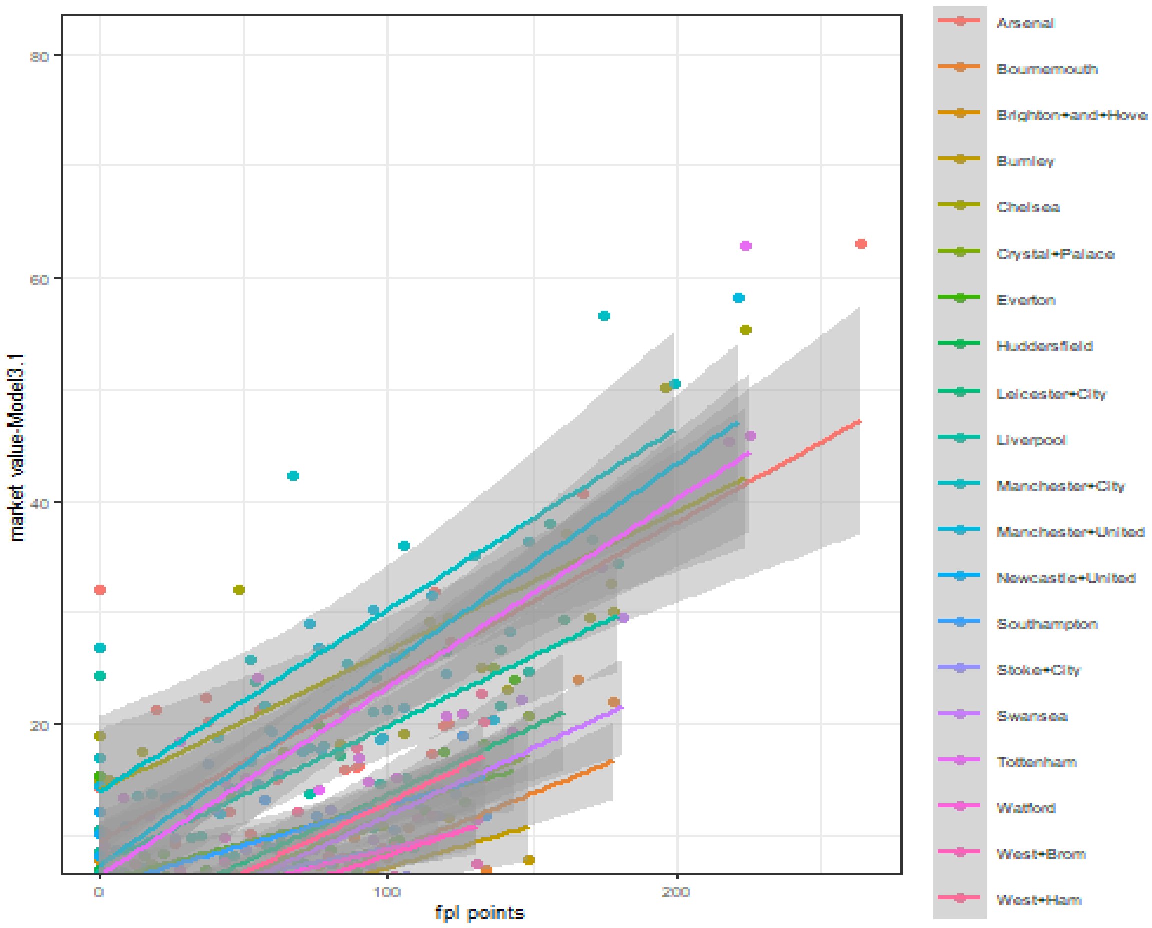

Figure 3

These are all level 2 work, then I need add big_club as level 3, model data again.

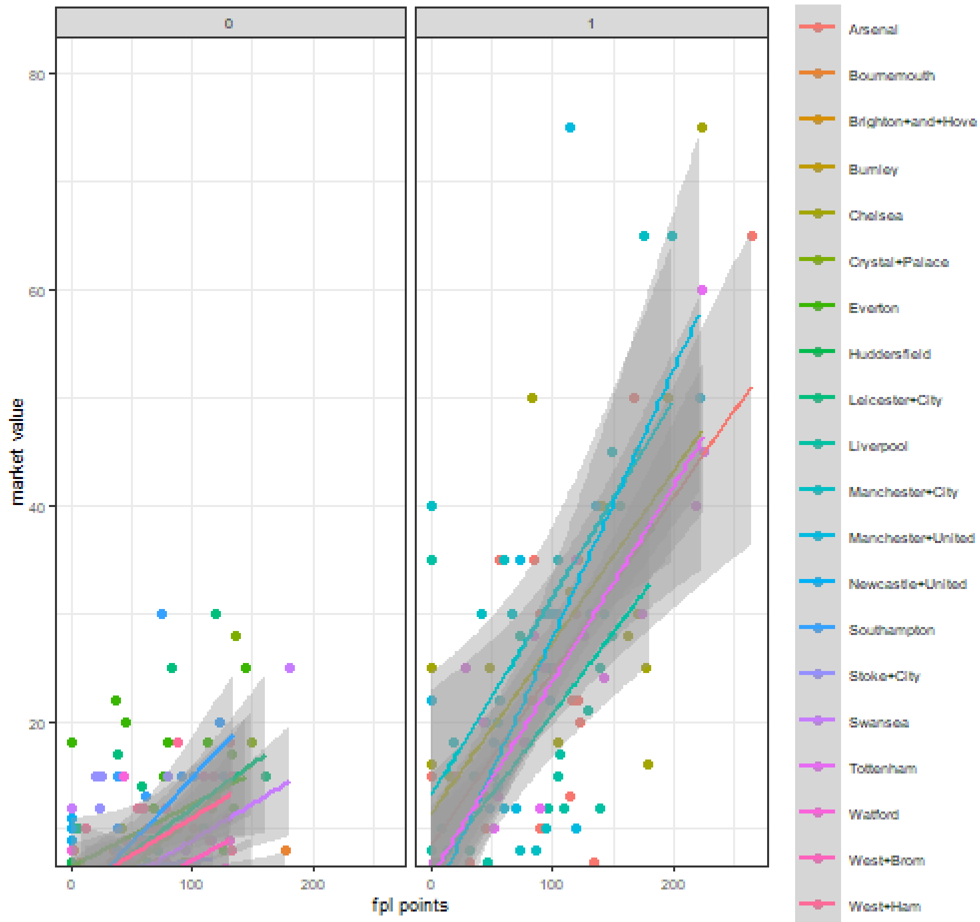

Figure 4

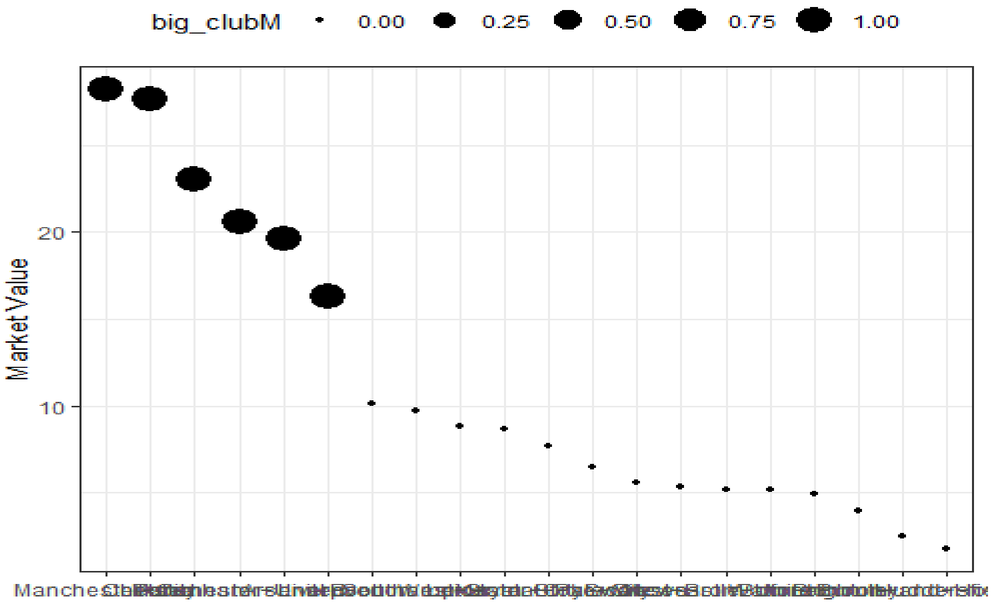

Figure 5

0 means normal club and 1 is big club. We can see that big club's player have very high value

than normal club players. I also calculate the total player market value for each club. In the English Premier League, Manchester City get highest value that is 564. Repetition what I did in level 2. In this time, with big_club add, the model become to market_value~1+(1|big_club) (market_value~fpl_points2+fpl_value2+(1|big_club) and (market_value~fpl_points2*fpl_value2+(1|big_club)+(1|big_club:club). The final plot is looking like this,

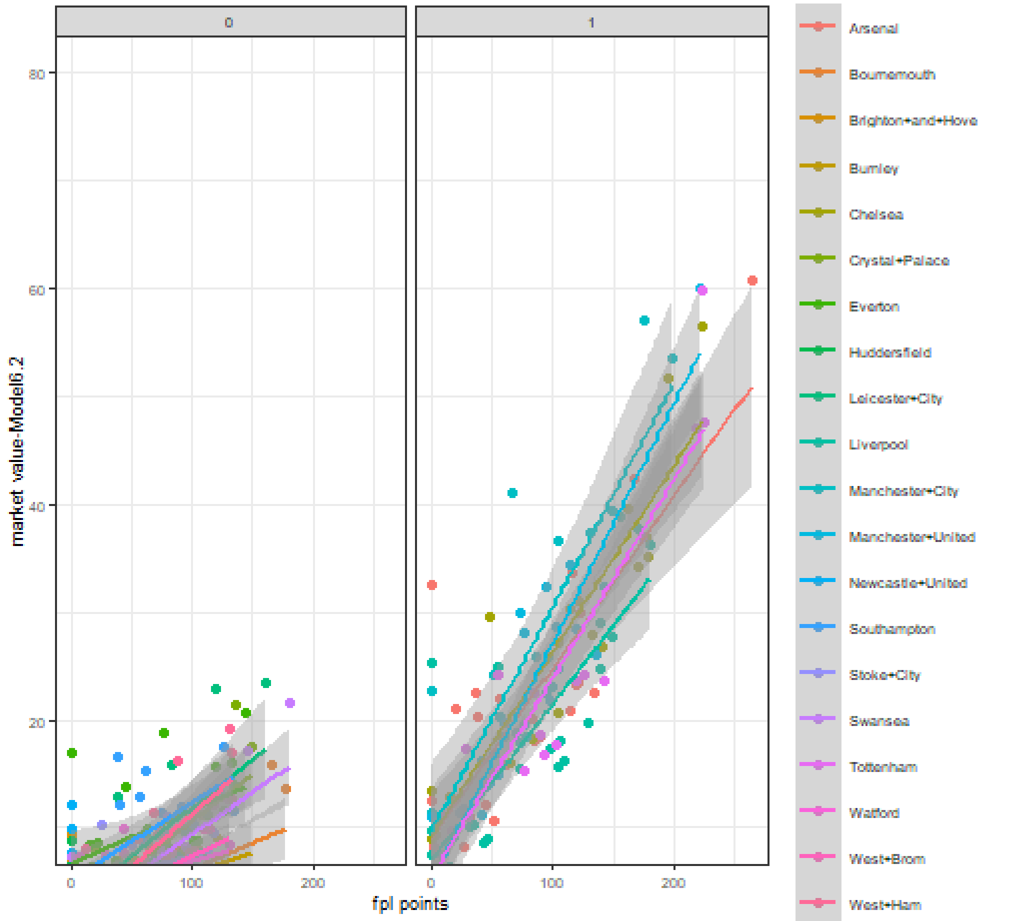

Figure 6

Result:

With model using and use anova to compare the model, we find that FPL value and point have relationship with player's market value. The reason for this may be the FPL point and value is made by player's performance and the player's performance will influence player's market value. For the big club, their player has much higher market value than normal club players, it maybe because the big club will buy player with higher price than other club, higher transfer fee will promote player's market value. The other reason maybe the big club have more attention, if the player has good performance, they will get more attention then promote their market value.

Discussion:

With my research, I got that big club player will have higher market value and it have relationship with FPL game. This dataset is only collected 2017-2018 season, if have dataset for last few years, I can get every club player's market value change through years. With that I can research that which player have good performance in these years and how each club strength changes with compare player's market value.

Appendix:

Level2 model

```
> model1 <- lmer(market_value~1+(1|club),REML=FALSE, data=epldata_final)
> summary(model1)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: market_value ~ 1 + (1 | club)
   Data: epldata_final

     AIC      BIC   logLik deviance df.resid
  3429.2   3441.5  -1711.6   3423.2      455

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.7724 -0.4411 -0.1137  0.2978  5.7527

Random effects:
 Groups   Name        Variance Std.Dev.
 club     (Intercept) 62.02    7.875
 Residual             91.29    9.554
Number of obs: 458, groups:  club, 20

Fixed effects:
            Estimate Std. Error t value
(Intercept)   11.061      1.818   6.085
> confint(model1)
Computing profile confidence intervals ...
               2.5 %   97.5 %
.sig01      5.790746 11.31377
.sigma      8.955001 10.22398
(Intercept) 7.319859 14.80295
> ICC(outcome="market_value", group="club", data=epldata_final)
[1] 0.4178369

> model2 <- lmer(market_value~fpl_points2+fpl_value2+(1|club),REML=FALSE, data=epldata_final)
> summary(model2)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: market_value ~ fpl_points2 + fpl_value2 + (1 | club)
   Data: epldata_final

     AIC      BIC   logLik deviance df.resid
  3035.0   3055.7  -1512.5   3025.0      453

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0070 -0.5076 -0.0517  0.4082  6.8860

Random effects:
 Groups   Name        Variance Std.Dev.
 club     (Intercept) 17.40    4.172
 Residual             38.93    6.239
Number of obs: 458, groups:  club, 20

Fixed effects:
             Estimate Std. Error t value
(Intercept) 11.024580   0.978085  11.272
fpl_points2  0.049037   0.007783   6.301
fpl_value2   4.813603   0.298428  16.130

Correlation of Fixed Effects:
            (Intr) fpl_p2
fpl_points2 -0.007
fpl_value2   0.004 -0.586
> confint(model2)
Computing profile confidence intervals ...
                2.5 %      97.5 %
.sig01      2.9933408  6.07876936
.sigma      5.8473498  6.67690570
(Intercept) 9.0074346 13.04235535
fpl_points2 0.0337506  0.06432395
fpl_value2  4.2205439  5.40947001
```

```
> model3 <- lmer(market_value~fpl_points2*fpl_value2+(1|club),REML=FALSE, data=epldata_final)
> summary(model3)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: market_value ~ fpl_points2 * fpl_value2 + (1 | club)
   Data: epldata_final

     AIC      BIC   logLik deviance df.resid
  3029.3   3054.0  -1508.6   3017.3      452

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0674 -0.4708 -0.0735  0.4232  7.0264

Random effects:
 Groups   Name        Variance Std.Dev.
 club     (Intercept) 17.99    4.241
 Residual             38.19    6.180
Number of obs: 458, groups:  club, 20

Fixed effects:
                        Estimate Std. Error t value
(Intercept)            10.673764   0.999987  10.674
fpl_points2             0.043086   0.008001   5.385
fpl_value2              4.399579   0.329469  13.354
fpl_points2:fpl_value2  0.007982   0.002849   2.802

Correlation of Fixed Effects:
            (Intr) fpl_p2 fpl_v2
fpl_points2  0.027
fpl_value2   0.059 -0.390
fpl_pnt2:_2 -0.125 -0.265 -0.441
> confint(model3)
Computing profile confidence intervals ...
                             2.5 %      97.5 %
.sig01                 3.050172121  6.17235667
.sigma                 5.791981382  6.61367422
(Intercept)            8.612030802 12.73143009
fpl_points2            0.027370646  0.05880520
fpl_value2             3.745652626  5.05655427
fpl_points2:fpl_value2 0.002381263  0.01357865
```

```
> anova(model2, model3)
Data: epldata_final
Models:
model2: market_value ~ fpl_points2 + fpl_value2 + (1 | club)
model3: market_value ~ fpl_points2 * fpl_value2 + (1 | club)
       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
model2    5 3035.1 3055.7 -1512.5   3025.1
model3    6 3029.3 3054.0 -1508.6   3017.3 7.7652  1   0.005326 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

```
> model3.2 <- lmer(market_value~fpl_points2*fpl_value2+(1+fpl_points|club),REML=FALSE, data=epldata_final)
boundary (singular) fit: see help('isSingular')
> summary(model3.2)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: market_value ~ fpl_points2 * fpl_value2 + (1 + fpl_points | club)
   Data: epldata_final

     AIC      BIC   logLik deviance df.resid
  2988.8   3021.8  -1486.4   2972.8      450

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2540 -0.4730 -0.0848  0.3740  6.8924

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 club     (Intercept)  2.249696 1.49990
          fpl_points   0.002165 0.04653  1.00
 Residual             34.425617 5.86733
Number of obs: 458, groups:  club, 20

Fixed effects:
                        Estimate Std. Error t value
(Intercept)            10.373884   0.995708  10.419
fpl_points2             0.039458   0.013651   2.891
fpl_value2              4.441555   0.307517  14.443
fpl_points2:fpl_value2 -0.001296   0.002764  -0.469

Correlation of Fixed Effects:
            (Intr) fpl_p2 fpl_v2
fpl_points2  0.803
fpl_value2   0.059 -0.210
fpl_pnt2:_2 -0.148 -0.168 -0.364
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Level3 model

```
> model4 <- lmer(market_value~1+(1|big_club)+(1|big_club:club),REML=FALSE, data=epldata_final)
> summary(model4)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: market_value ~ 1 + (1 | big_club) + (1 | big_club:club)
   Data: epldata_final

     AIC      BIC   logLik deviance df.resid
  3403.2   3419.7  -1697.6   3395.2      454

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.6647 -0.4243 -0.1185  0.2961  5.6426

Random effects:
 Groups        Name        Variance Std.Dev.
 big_club:club (Intercept)  7.048   2.655
 big_club      (Intercept) 64.488   8.030
 Residual                  91.238   9.552
Number of obs: 458, groups:  big_club:club, 20; big_club, 2

Fixed effects:
            Estimate Std. Error t value
(Intercept)   14.193      5.736   2.474
```

```
> model5 <- lmer(market_value~fpl_points2+fpl_value2+(1|big_club)+(1|big_club:club),REML=FALSE, data=epldata_final)
> summary(model5)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: market_value ~ fpl_points2 + fpl_value2 + (1 | big_club) + (1 |      big_club:club)
   Data: epldata_final

     AIC      BIC   logLik deviance df.resid
  3014.2   3039.0  -1501.1   3002.2      452

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0630 -0.4692 -0.0322  0.4378  6.8693

Random effects:
 Groups        Name        Variance Std.Dev.
 big_club:club (Intercept)  2.558   1.599
 big_club      (Intercept) 17.775   4.216
 Residual                  38.908   6.238
Number of obs: 458, groups:  big_club:club, 20; big_club, 2

Fixed effects:
             Estimate Std. Error t value
(Intercept) 12.665342   3.023851   4.188
fpl_points2  0.047189   0.007478   6.310
fpl_value2   4.770814   0.296881  16.070

Correlation of Fixed Effects:
            (Intr) fpl_p2
fpl_points2 -0.003
fpl_value2  -0.014 -0.577
>
> model6 <- lmer(market_value~fpl_points2*fpl_value2+(1|big_club)+(1|big_club:club),REML=FALSE, data=epldata_final)
> summary(model6)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: market_value ~ fpl_points2 * fpl_value2 + (1 | big_club) + (1 |      big_club:club)
   Data: epldata_final

     AIC      BIC   logLik deviance df.resid
  3008.0   3036.9  -1497.0   2994.0      451

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1290 -0.4718 -0.0493  0.4047  7.0085

Random effects:
 Groups        Name        Variance Std.Dev.
 big_club:club (Intercept)  2.611   1.616
 big_club      (Intercept) 18.409   4.291
 Residual                  38.169   6.178
Number of obs: 458, groups:  big_club:club, 20; big_club, 2

Fixed effects:
                        Estimate Std. Error t value
(Intercept)            12.338660   3.078129   4.008
fpl_points2             0.042011   0.007632   5.505
fpl_value2              4.343105   0.328898  13.205
fpl_points2:fpl_value2  0.008096   0.002805   2.887

Correlation of Fixed Effects:
            (Intr) fpl_p2 fpl_v2
fpl_points2  0.006
fpl_value2   0.004 -0.396
fpl_pnt2:_2 -0.037 -0.237 -0.447
> anova(model5,model6)
Data: epldata_final
Models:
model5: market_value ~ fpl_points2 + fpl_value2 + (1 | big_club) + (1 | big_club:club)
model6: market_value ~ fpl_points2 * fpl_value2 + (1 | big_club) + (1 | big_club:club)
       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
model5    6 3014.2 3039.0 -1501.1   3002.2
model6    7 3008.0 3036.9 -1497.0   2994.0 8.2498  1   0.004076 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

```
> summary(model6.2)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: market_value ~ fpl_points2 * fpl_value2 + (1 + fpl_points2 |      big_club) + (1 + fpl_points2 | big_club:club)
   Data: epldata_final

     AIC      BIC   logLik deviance df.resid
  2997.2   3042.5  -1487.6   2975.2      447

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2670 -0.4327 -0.1006  0.3624  6.9395

Random effects:
 Groups        Name        Variance  Std.Dev. Corr
 big_club:club (Intercept) 3.102e+00 1.76111
               fpl_points2 3.279e-04 0.01811  0.97
 big_club      (Intercept) 3.004e+01 5.48044
               fpl_points2 1.791e+01 4.23238  -0.06
 Residual                  3.420e+01 5.84811
Number of obs: 458, groups:  big_club:club, 20; big_club, 2

Fixed effects:
                        Estimate Std. Error t value
(Intercept)            12.021804   3.913048   3.072
fpl_points2             0.056646   2.992755   0.019
fpl_value2              4.374485   0.311161  14.059
fpl_points2:fpl_value2 -0.001607   0.003072  -0.523

Correlation of Fixed Effects:
            (Intr) fpl_p2 fpl_v2
fpl_points2 -0.055
fpl_value2   0.003 -0.001
fpl_pnt2:_2 -0.020 -0.001 -0.397
optimizer (nloptwrap) convergence code: 0 (OK)
unable to evaluate scaled gradient
Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
```

CODE:

```
library(lme4)
library(readr)
library(ggplot2)
library(merTools)
library(dplyr)
epldata_final <- read_csv("C:/Users/Haoyue/Desktop/epldata_final.csv")
head(epldata_final)


ggplot(data=epldata_final, aes(x=name,y=market_value)) +
   geom_point() +
   geom_smooth(method = "lm", se = TRUE) +
   xlab("name") + ylab("value") +
   facet_wrap( ~ club, ncol=5)

#level2
epldata_final$fpl_points2 <- scale(epldata_final$fpl_points,scale = F)
epldata_final$fpl_value2 <- scale(epldata_final$fpl_value,scale = F)



model1 <- lmer(market_value~1+(1|club),REML=FALSE, data=epldata_final)
summary(model1)
confint(model1)
ICC(outcome="market_value", group="club", data=epldata_final)
```

```
model2 <- lmer(market_value~fpl_points2+fpl_value2+(1|club),REML=FALSE,
data=epldata_final)
summary(model2)
confint(model2)
model3 <- lmer(market_value~fpl_points2*fpl_value2+(1|club),REML=FALSE,
data=epldata_final)
summary(model3)
confint(model3)
anova(model2, model3)
epldata_final$model3.1<-predict(model3, newdata=epldata_final)
theme_set(theme_bw(base_size = 7, base_family = ""))

ggplot(data = epldata_final, aes(x = fpl_points, y=market_value,group=club))+
  coord_cartesian(ylim=c(10,80))+
  geom_point(aes(colour = club))+
  geom_smooth(method = "lm", se = TRUE,aes(colour = club))+
  xlab("fpl points")+ylab("market value")
ggplot(data = epldata_final, aes(x = fpl_points, y=model3.1,group=club))+
  coord_cartesian(ylim=c(10,80))+
  geom_point(aes(colour = club))+
  geom_smooth(method = "lm", se = TRUE,aes(colour = club))+
  xlab("fpl points")+ylab("market value-Model3.1")

model3.2 <- lmer(market_value~fpl_points2*fpl_value2+(1+fpl_points|club),REML=FALSE,
data=epldata_final)
summary(model3.2)
#level3
ggplot(data = epldata_final, aes(x = fpl_points, y=market_value,group=club))+
  facet_grid(~big_club)+
  coord_cartesian(ylim=c(10,80))+
  geom_point(aes(colour = club))+
  geom_smooth(method = "lm", se = TRUE,aes(colour = club))+
  xlab("fpl points")+ylab("market value")


Plot.Means<-epldata_final %>% group_by(club) %>%
  dplyr::summarize(marketM=mean(market_value, na.rm=TRUE),
                   big_clubM=mean(big_club, na.rm=TRUE))

market_with_club<-ggplot(data = Plot.Means, aes(x = reorder(club, -marketM),
y=marketM))+
  geom_point(aes(size = big_clubM))+
  xlab("")+ylab("Market Value")+
  theme_bw()+
```

```
  theme(legend.position = "top")
market_with_club

aggregate(epldata_final$market_value, list(epldata_final$club), FUN=sum)


model4 <- lmer(market_value~1+(1|big_club)+(1|big_club:club),REML=FALSE,
data=epldata_final)
summary(model4)
model5 <-
lmer(market_value~fpl_points2+fpl_value2+(1|big_club)+(1|big_club:club),REML=FALSE,
data=epldata_final)
summary(model5)
model6 <-
lmer(market_value~fpl_points2*fpl_value2+(1|big_club)+(1|big_club:club),REML=FALSE,
data=epldata_final)
summary(model6)
anova(model5,model6)
model6.2 <-
lmer(market_value~fpl_points2*fpl_value2+(1+fpl_points2|big_club)+(1+fpl_points2|big_cl
ub:club),REML=FALSE, data=epldata_final)
summary(model6.2)

epldata_final$model6.2<-predict(model6.2, newdata=epldata_final)
theme_set(theme_bw(base_size = 7, base_family = ""))


ggplot(data = epldata_final, aes(x = fpl_points, y=model6.2,group=club))+
  facet_grid(~big_club) +
  coord_cartesian(ylim=c(10,80))+
  geom_point(aes(colour = club))+
  geom_smooth(method = "lm", se = TRUE,aes(colour = club))+
  xlab("fpl points")+ylab("market value-Model6.2")
```