**Traditional Statistics And Scoring In The NBA**

**Introduction:**

What truly drives offensive prowess in the NBA? How many points can a team score in 48 minutes? Common knowledge in 2015 is that Western Conference teams score more points than Eastern Conference teams. We compare conferences and investigate traditional NBA statistics along with a few advanced NBA statistics to determine how an NBA team can score more points.

Average per game statistics were retrieved by team via python from SportsRadar API for the past four NBA seasons. These data range from offensive rebounds, defensive rebounds, to three pointers, turnovers and field goal efficiency. In total, 31 different statistics were examined. These data are completely observational, thus no causal association can be inferred.

**Analysis:**

In order to preliminarily research linear associations with points a correlation matrix is ran, with the following results:

|  |  |
| --- | --- |
| Variable | Correlation Coefficient |
| Three Point Efficiency | 0.66 |
| Defensive Rebounds | 0.50 |
| CONFERENCE (Two Factors – East & West) | 0.47 |
| Assists | 0.47 |
| Game Pace | 0.39 |
| Personal Fouls Drawn | 0.38 |

Immediately, three point efficiency (three points made \* three point percentage) tops the list, followed by defensive rebounds per game, the team’s conference, assists per game, game pace (possessions per 48 minutes) and personal fouls drawn.

Personal fouls drawn is obviously correlated with points based on Pearson’s R. However, is there a visual linear correlation?

|  |  |
| --- | --- |
| Scatter Plot Matrix of Points and Personal Fouls Drawn  Showing obvious linear relationship | Macintosh HD:Users:patrickcorynichols:Desktop:Screen Shot 2015-08-13 at 8.15.00 PM.png |

A scatter plot matrix shows linear relationships for all of the highly correlated variables. Therefore, a rich model is fit with a total N of 120. This model is :

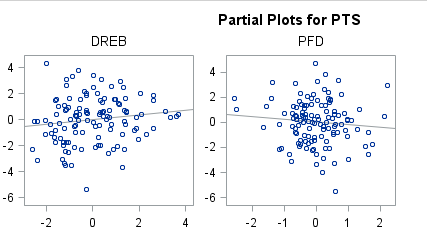
{pts| threepe, dreb, CONFERENCE, assists, game pace, pfd}.

The model fit results in the following summary:

|  |  |
| --- | --- |
| Initial fit shows an r-sqared of 0.8484 with an adjusted r-sqaured of 0.8390. All variables in the model are statistically significant except for defensive rebrounds and personal fouls drawn with relatively acceptable variance inflation. | Macintosh HD:Users:patrickcorynichols:Desktop:Screen Shot 2015-08-13 at 8.21.46 PM.png |

The overall F for the initial fit is extremely significant at F of 89.56 and p <.0001. The model is able to explain 2021 of 2382 sum of squared error when compared to the equal means reduced model.

Because personal fouls drawn and defensive rebounds are statistically insignificant in this model, it may be wise to remove them. However, we first investigate partial residuals.

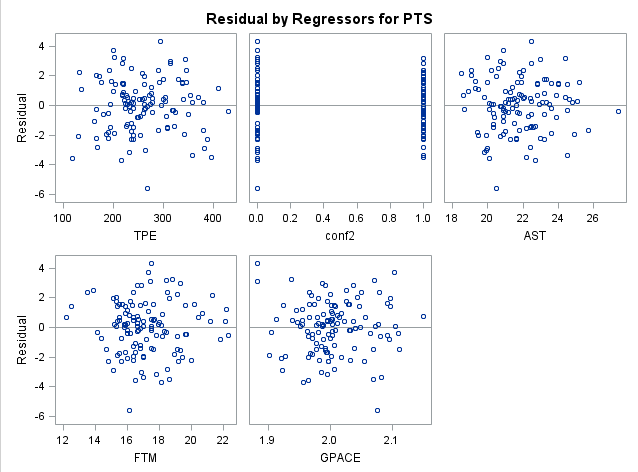


Both defensive rebounds and personal fouls drawn have a relatively linear relationship when accounting for all other variable effects in the model. That being said, personal fouls drawn is not linear enough to stay in the model. We will move to another model without personal fouls drawn and re-fit the model as:

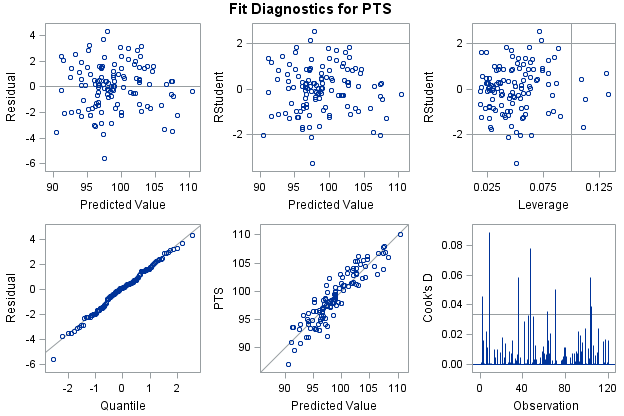
{pts | threepe CONFERENCE, assists, game pace}.

Based on the new model excluding defensive rebounds and personal fouls drawn, the overall significant is still p < .0001 with an even higher F-value of 122.69. To further confirm both defensive rebound and personal fouls drawn variables can be dropped, an extra sum of squares F-test is ran. The result is an F-statistic of 1.86 on 2 and 112 degrees of freedom. The resultant p value is 0.16, indicating there is no evidence that defensive rebounds and personal fouls drawn account for an increase in points after accounting for three point efficiency, conference, assists, free throws made and game pace. Therefore, we move forward with our reduced model and check residual plots to ensure assumptions are met for multiple linear regression.

The following normal residual plots are produced:

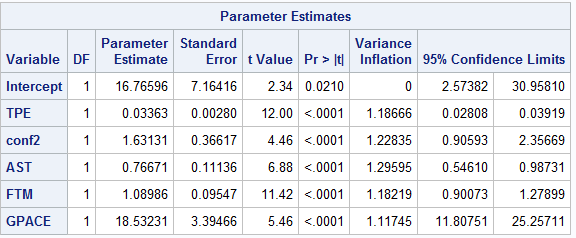


No transformations are necessary in this case. A model r squared of 0.84 is still maintained even with our reduced model without defensive rebounds and personal fouls drawn. This means that our new model is able to explain 84% of the variance in our response, points. Individual variable residuals as well as full model residuals look acceptable:



Influential points are minimal with studentized residuals showing acceptable standard deviations for the model with 4/120 observations being more than +- 2 standard deviations away from the mean. Normality assumptions are met with ease per a qq plot.

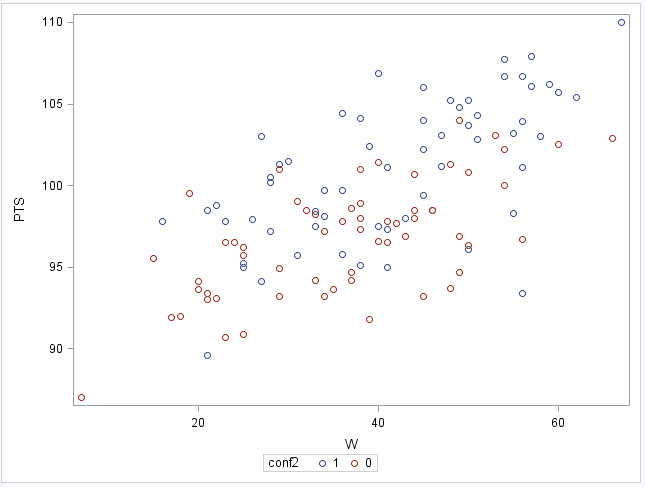
Does the Western Conference score more points than the Easter Conference? Our conference variable, where 0 = Eastern and 1 = indicates, all other variables held constant, that Western Conference teams score 1.6 more points per game than Eastern Conference teams with a confidence interval of 0.91 to 2.36 points:



In the proposed parallel lines model, Western Conference teams do indeed score more points than Eastern Conference teams. The extra 1.63 points in the model represent a positive increase in the intercept of the model when considering a Western Conference team compared to the reference Eastern Conference team.

**Conclusion:**

So what statistics drive prolific scoring in the NBA? Based on our model, there is an association between points, three point effectiveness, assists, free throws made and game pace. Further, there is evidence to show that Western conference teams score more than Eastern Conference teams. Additional analysis into why this association is significant is warranted as a next step in our analysis. However, as a first step into this analysis, it is evident from a scatter plot on points and wins, that prolific scoring teams may be more successful than lower scoring teams:



SAS:

**PROC** **IMPORT** OUT = NBA

DATAFILE = '\\Client\C$\Users\patrickcorynichols\Desktop\nba2.csv'

DBMS = CSV REPLACE;

GETNAMES = YES;

DATAROW = **2**;

**RUN**;

**DATA** NBA2;

SET NBA;

TWOPM = FGM-TPM;

TWOPA = FGA-TPA;

TWOPCT = TWOPM/TWOPA;

TWOPE = TWOPM\*TWOPCT;

TPE = TPM\*TPCT;

FGE = FGM\*FGPCT;

int1 = AST\*TPE;

GPACE = (FGA+TOV)/MIN;

IF conf = 'W' THEN conf2 = **1**; ELSE conf2= **0**;

**RUN**;

**PROC** **CORR** DATA = NBA2;

VAR PTS AST DREB OREB STL FTM TPE TWOPE FGE CONF2 GPACE PFD;

**RUN**;

**PROC** **SGSCATTER** DATA = NBA2;

MATRIX PTS AST DREB OREB STL / GROUP = conf2;

**RUN**;

Symbol1 color=blue interpol=join line=**1** value=dot;

Symbol2 color=red interpol=join line=**2** value=star;

**PROC** **SGSCATTER** DATA = NBA2;

PLOT PTS\*W / GROUP = conf2;

LEGEND across = **1** down = **2**;

**RUN**;

**PROC** **GPLOT** DATA = NBA2;

PLOT PTS\*W / group = conf2;

**RUN**;

**PROC** **SGSCATTER** DATA = NBA2;

MATRIX PTS PFD;

**RUN**;

**PROC** **REG** DATA = NBA2;

MODEL PTS = TPE CONF2 AST FTM GPACE / CLB CLM CLI R VIF;

**RUN**;

**PROC** **REG** DATA = NBA2;

MODEL PTS = TPE CONF2 AST FTM GPACE DREB PFD / VIF;

**RUN**;

**QUIT**;