Does Defense Win Championships?

Arthur Lennard

Introduction and Methods

It is a fairly common cliche in sports these days that "offense wins games, defense wins championships." This is just an old saying however, and perhaps the reasoning of the more defensive-minded of the athletes that do not feel they get the same recognition as those responsible for all the scoring. Curious as to whether they might have a point though, I have decided to run various multiple regression tests to find out. Using data from Andrew Sundberg's college basketball data set via Kaggle, I will gather values from last year's college basketball season in the following categories: adjusted offensive efficiency, adjusted defensive efficiency, effective field goal percentage shot, effective field goal percentage allowed, and seed in the NCAA tournament. Adjusted offensive efficiency refers to a team's average points scored per 100 possessions, while adjusted defensive efficiency is the average points allowed per 100 possessions. By using these values as factors in a multiple regression with tournament seed as the dependent variable, I will be able to determine which has a greater impact on the deviation in seed, and therefore which value is a better predictor of what seed a team will be in the March tournament. Similarly, I will carry out the same process with effective field goal percentage shot and allowed, to determine which of these has a greater impact on a team's seed as well. Both of these will provide evidence toward whether defensive or offensive qualities carry more weight when determining how good a team is considered to be. Additionally, since lower seeds should theoretically have an easier path to the final of the tournament in respect to the seeds they are set to play against, I will be able to provide evidence for if offense or defense play a greater role in providing a team with a better chance to get to the championship game. I will be selecting each of the 68 teams that qualified for the 2021 NCAA Men's Basketball Tournament for this analysis, and comparing their offensive and defensive stats to their seed in the tournament. P-values in this study to determine significance of the relationships in the regression tests will be compared to $\alpha = 0.05$.

Results

```
Call:
lm(formula = seed ~ effDef + effOff)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.8306 -1.4905 -0.1311  1.5421  3.6113 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 13.56944    8.10331   1.675   0.0989 .  
effDef       0.48691    0.05731   8.496 4.32e-12 ***
effOff      -0.46599    0.03903 -11.938  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.926 on 64 degrees of freedom
Multiple R-squared:  0.8396,    Adjusted R-squared:  0.8346 
F-statistic: 167.5 on 2 and 64 DF,  p-value: < 2.2e-16
```

Figure A

Figure A depicts the output from the multiple regression test comparing adjusted offensive and defensive efficiency to seed. First of all, although the p-value associated with the intercept of the resulting regression equation is not significant compared to the $\alpha = 0.05$ level, each of the p-values connected to the slope of the graph from the efficiency variables are. This means that there is a significant impact on tournament seed by offensive and defensive efficiency. Additionally, the multiple r-squared tells us that 83.96% of the variation in seed can be explained by these variables in this test, which is reasonably high. In this output, the offensive efficiency coefficient is about -0.466. It makes sense that this value should be inversely proportional to seed, as teams that are more efficient on offense should theoretically earn lower seeds in the tournament. What is peculiar however, is that the coefficient for defensive efficiency is a positive value, indicating that defensive efficiency is directly proportional to seed.

```
Call:
lm(formula = seed ~ shootD + shootO)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.2379 -3.6553 -0.0013  3.9041  8.1861 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  20.7452    18.2739   1.135   0.2605  
shootD        0.2876     0.3154   0.912   0.3653  
shootO       -0.4916     0.2049  -2.400   0.0193 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.58 on 64 degrees of freedom
Multiple R-squared:  0.09219,	Adjusted R-squared:  0.06382 
F-statistic:  3.25 on 2 and 64 DF,  p-value: 0.04527
```

Figure B

Figure B displays the results of the same multiple regression test from Figure A, but with effective shooting percentage for offense and defense. Right away, this output indicates that there is not a significant relationship between shooting efficiency on either side of the ball and seed in the NCAA tournament. This is because each of the p-values for the variables and the intercept are all higher than the selected alpha value of 0.05. Moreover, the multiple r-squared value for this test suggests that less than 10% of the variation in seed is due to either of these variables.

```
> lm.beta(model_3)
   effDef     effOff
 0.4575734 -0.6429478
> lm.beta(model_4)
   shootD     shootO
 0.1086098 -0.2858545
```

Figure 3

Finally, figure 3 displays the post hoc test for the standardized regression coefficient of each of the multiple regressions. While the values for the shooting efficiency test (model 4) are not relevant, the results for offensive and defensive efficiency (model 3) do reveal valid information. Since the absolute value of the offensive statistic is greater than the defensive one, we can conclude that a team's offensive efficiency has a greater impact on the relationship with the team's seed than their defensive efficiency does.

Conclusion and Discussion

The results of this study do not indicate that defense wins championships. While both adjusted offensive and defensive efficiency ratings proved to be equitable predictors of a team's seed in the NCAA tournament, the relationship that seed has with defensive efficiency indicates that being worse at defensive is more closely associated with being a lower seed. Conversely, the results of the study indicate that the higher a team's offensive efficiency rating is, the lower a seed they tend to be. Additionally, offensive efficiency has a stronger impact on seed than defensive efficiency, indicating that the offensive statistic is more relevant to predicting what seed a team will be. As for effective shooting percentage both for and against a team, neither indicated a significant relationship is respect to seed in the tournament. This study, therefore, does not provide any evidence that playing effective defense puts men's college basketball teams in better positions to win championships come March. In the future, more research should be done on additional past seasons of college basketball, and also done taking into account more variety in defensive statistics.