Conor Juengst
Sarah Winfield
Prof. Lin
STAT 520

Factor Analysis:

Opposition Statistics in NCAA Division 1 Men's Basketball

Introduction and Data:

Our objective for this project was to identify a few discrete factors related to a team's defensive performance. Specifically, we were looking to find a relatively small number of unobservable random variables $F_1...F_m$ that would both explain a significant proportion of the variance and reduce the dimensionality of our data set.

The data we chose for our analysis was the '2014-15 Advanced Opponent Stats' data set from College Basketball Reference.[1] This particular data set recorded the opposition statistics[2] from the 2014-2015 season for all 351 Division 1 NCAA Men's basketball programs, and included possession-based metrics as well as other non-traditional statistics.

We focused on ten of the statistical categories in our data set for the purpose of this analysis. Those categories were: offensive rating, effective field goal percentage, true shooting percentage, free throw attempt rate, free throws per field goal attempt, turnover percentage, three point attempt rate, assist percentage, offensive rebounding percentage, and steal percentage. We chose to exclude the block percentage, pace of play, and total rebounding percentage statistics from our analysis for differing reasons. Block percentage was excluded because a block does not always result in a change of possession, and therefore does not always benefit the defense to the same extent. We removed pace of play after we determined that there is not a significant correlation between pace of play and the success of the defense. Lastly, total rebounding percentage was dropped from our model because it includes defensive rebounding in its calculation, which is an element that does not reflect a team's defensive performance.

Hypothesis and Assumptions:

---

[1] <*http://www.sports-reference.com/cbb/seasons/2015-advanced-opponent-stats.html*>

[2] Opposition Statistics - the statistics accumulated by the opponents of a particular school
ex. Virginia's ORtg of 87.2 indicates that, on average, their opponents score 87.2 pts per 100 poss.

We elected to use an orthogonal factor model for the purpose of our analysis which can be expressed as: $x - \mu = LF + \varepsilon$. Furthermore, we needed to make a few assumptions regarding $F_1...F_m$, our *common factors*, and $\varepsilon_1...\varepsilon_p$, our error terms, or *specific factors*. First, we had to assume that F and ε are independent since we are employing an orthogonal model. Additionally, we needed to assume that the expectation of F and ε is 0 in each case, and that the variance of F is given by the identity matrix **I** while the variance of ε is conveyed by covariance matrix $\mathbf{\Psi}$. Finally, if using the maximum likelihood method for our factor model we have to assume the data is from a normal distribution ($X \sim N_p$).

Factor Model:

Ultimately, we chose to use the principal component method for our factor analysis rather than the maximum likelihood method. Our reasoning for this decision had to do with our factors and their interpretation, as well as consideration of the total communality for each method.[3] When testing ML models, it was inconclusive as to how many factors would be appropriate for our model. We generated three, four, and five factor ML models, each of which offered a distinct yet similar interpretation of the factor structure. In contrast, our principal component model was much less ambiguous in regards to how many factors would be optimal, and offered a higher total communality which aided us in our decision to use a PC model.

Factor Interpretation:

Factor 1: Efficiency of the shooting defense.

Offensive rating, effective field goal percentage, and true shooting percentage (ORtg, eFG%, TS%)

These three statistics all take into account how well the defense is affecting shooting. If the defense is effectively limiting the offense's shooting, their effective field goal percentage and true shooting percentage will be lower, and their offensive rating will reflect the diminished efficiency. This factor can be thought of as the overall efficiency of a team's shooting defense.

Factor 2: Level and precision of defensive aggression.

Free throw attempt rate, free throws per field goal attempt, and turnover percentage (FTAr, FT/FGA, TOV%)

[3] Total communality is a measure of the variance explained by our factors

When the defense is playing aggressively it can result in outcomes from two opposing sides of the spectrum. If the aggression is precise, it will result in turnovers and possession changes without the accumulation of unnecessary fouls. On the contrary, if the aggression is lacking precision, it will result in a higher rate of personal fouls which will then lead to foul trouble for the defense and an efficient source of points for the offense. There also exists a relation between the free throw metrics and TOV% as teams that are more aggressive, and attack the basket more, will experience a higher volume of turnovers and free throws attempted in comparison to less assertive teams.

Factor 3: Rate at which the opponent creates open shots, and in particular, 3-pointers.

Three-point attempt rate and assist percentage (3PAr, AST%)

These two categories are grouped together in factor 3 due to the nature of three-point shots. It has been shown that over 80% of three-point field goals are assisted at the NBA level[4], and while collegiate basketball may not have such an extreme relation between these rates, it is evident that there exists a strong correlation between the two. This is because it is difficult for a player to create a high-efficiency three-point shot on their own. Instead, effective long-range field goal attempts are typically generated by absorbing the defense's attention before passing the ball to a vacated shooter on the perimeter. From a defensive view, limiting incisive passing and committing less resources to help defense results in a lower number of assists and can reduce the rate at which an offense attempts three-point shots.

Factor 4: Rate at which the opponent creates new possession opportunities.

Offensive rebounding percentage and steal percentage (ORB%, STL%)

Along with limiting the effectiveness of an opponent's shooting, successful defenses also restrict the total number of possessions that an opposing offense gets. If the opposition gathers an offensive rebound, they are effectively extending the current possession. On the other hand, when the opposition steals the ball they are gaining a possession outright. Ideally, a defense aims to control the number of extra possessions a team gains through offensive rebounds and steals.

Principal Component Comparison:

The last part of our project was a principal component breakdown that compares the defensive performance of San Diego State University men's basketball team against other well known basketball programs across the country. Initially, we began by contrasting the simple

[4] Kirk Goldsberry <*http://grantland.com/the-triangle/the-rainmakers-players-who-create-the-most-corner-3s/*> "As a whole, 84 percent of NBA 3s are assisted [...] "

statistics that corresponded to each of our factors. We found that SDSU's defense excelled nationally in opponent offensive rating, free throw rate, assist percentage, and turnover rate. For instance, the average opponent offensive rating for the Aztecs was 13.58 points lower than the national average this season. This means that San Diego State held its opponents to an average of 87.7 points per 100 possessions or .877 points per possession compared to the 1.01 point national average. San Diego State also compared favorably in free throw attempt rate where the national average was 11.8% higher than the Aztecs'. In this case, opponents only made one free throw for every four field goals attempted. SDSU also held the opposition to an assist rate almost 7% lower than the national average, and caused their opponents to turn the ball over at a rate 1.75% above the mean.

We used principal component 1 (PC1) for the purpose of our comparison as it measured the overall performance of a team's shooting defense and accounted for the largest proportion of variance. After using standardized results to solve for a PC1 total for each school, we ranked the defense of the schools that we chose. A lower PC1 result indicated a higher rating of that team's defense, with Kentucky's score of 43.833 being the best in the country. We selected five schools to rank alongside SDSU which were: Kentucky, Duke, Arizona, New Mexico, and UNLV. The results were not alarming, as San Diego State University is known nationally for the way that its defense restricts an opponent's offense, as indicated by its low PC1 score of 47.970. Rating better than both Duke and the University of Arizona, this PC1 test shows that at least one of the areas that San Diego State triumphs in against these two teams is defense where they also hold a significant edge over their conference rivals, New Mexico and UNLV.

Conclusion:

For this project we successfully identified four underlying factors that define a successful college basketball defense. Our final factor model, using the principal component method with four factors, explained 81.87% of the variation and yielded an interpretable factor structure.