



SAS® GLOBAL FORUM 2016



IMAGINE. CREATE. INNOVATE.

Using Parametric and Nonparametric
Tests to Assess the Decision of the
NBA's 2014-2015 MVP Award

#SASGF



Using Parametric and Nonparametric Tests to Assess the Decision of the NBA's 2014-2015 MVP Award

Sherrie Rodriguez, MS in Applied Statistics

Advising Faculty: Dr. Bradley Barney

Kennesaw State University

Abstract

Stephen Curry, James Harden, and LeBron James are considered to be three of the most gifted professional basketball players in the National Basketball Association (NBA). Each year the KIA Most Valuable Player (MVP) award is given to the best player in the league. Stephen Curry currently holds this title, followed by James Harden and LeBron James, the first two runners-up. The decision for MVP was made by a panel of judges comprised of 129 sportswriters and broadcasters, along with fans who were able to cast their votes through NBA.com. *Did the judges make the correct decision? Is there statistical evidence that indicates that Stephen Curry is indeed deserving of this prestigious title over James Harden and LeBron James? Is there a significant difference between the two runners-up?* These are some of the questions that are addressed through this project.

Using data collected from NBA.com for the 2014-2015 season, a variety of parametric and nonparametric k-sample methods were used to test 20 quantitative variables. In an effort to determine which of the three players is the most deserving of the MVP title, post-hoc comparisons were also conducted on the variables that were shown to be significant. The time-dependent variables were standardized, as there was a significant difference in the number of minutes each athlete played. These variables were then tested and compared with those that had not been standardized. This led to significantly different outcomes, indicating that the results of the tests could be misleading if the time variable is not taken into consideration. Using the standardized variables, the results of the analyses indicate that there is a statistically significant difference in the overall performance of the three athletes, with Stephen Curry outplaying the other two players. However, the difference between James Harden and LeBron James is not so clear.

Additional analyses were conducted using the NBA.com data collected from the past six seasons. These were the years that the three contenders had in common in the NBA, and the variables were the same as those used in the previous analyses. In an attempt to visualize the differences in the players' overall performances, the variables that were found to be significant were then used to plot the player separation through multivariate methods.

Objectives

- I. The main objective of this project is to determine whether or not there is statistical evidence to support the decision of the NBA's 2014-2015 Award. The steps towards this objective are the following:
 - Determine if there are significant global differences between the three players
 - If there are global differences, determine if there are significant paired differences
 - Compare and contrast analyses using standardized vs unstandardized time-dependent variables
- II. Additional objectives using data from the past six seasons:
 - Apply multivariate methods in order to describe player separation
 - Visualize player separation



Using Parametric and Nonparametric Tests to Assess the Decision of the NBA's 2014-2015 MVP Award

Sherrie Rodriguez, MS in Applied Statistics

Advising Faculty: Dr. Bradley Barney

Kennesaw State University

Outline of Methods

I. Data Preparation

- Creation of new variables
- Standardization of all time-dependent variables

II. K-Sample Methods

- ANOVA using PROC GLM
- Kruskal-Wallis Test using PROC NPAR1WAY
- Comparison of parametric vs. nonparametric methods

III. Post-Hoc Comparisons

- Fisher's Least-Significant Difference (LSD) Test
- Wilcoxon Rank-Sum Test

IV. Comparing Standardized vs. Unstandardized Variables

V. Multivariate Methods

- Discriminant analysis using PROC CANDISC
- Contour plot using PROC KDE

I. Data Preparation

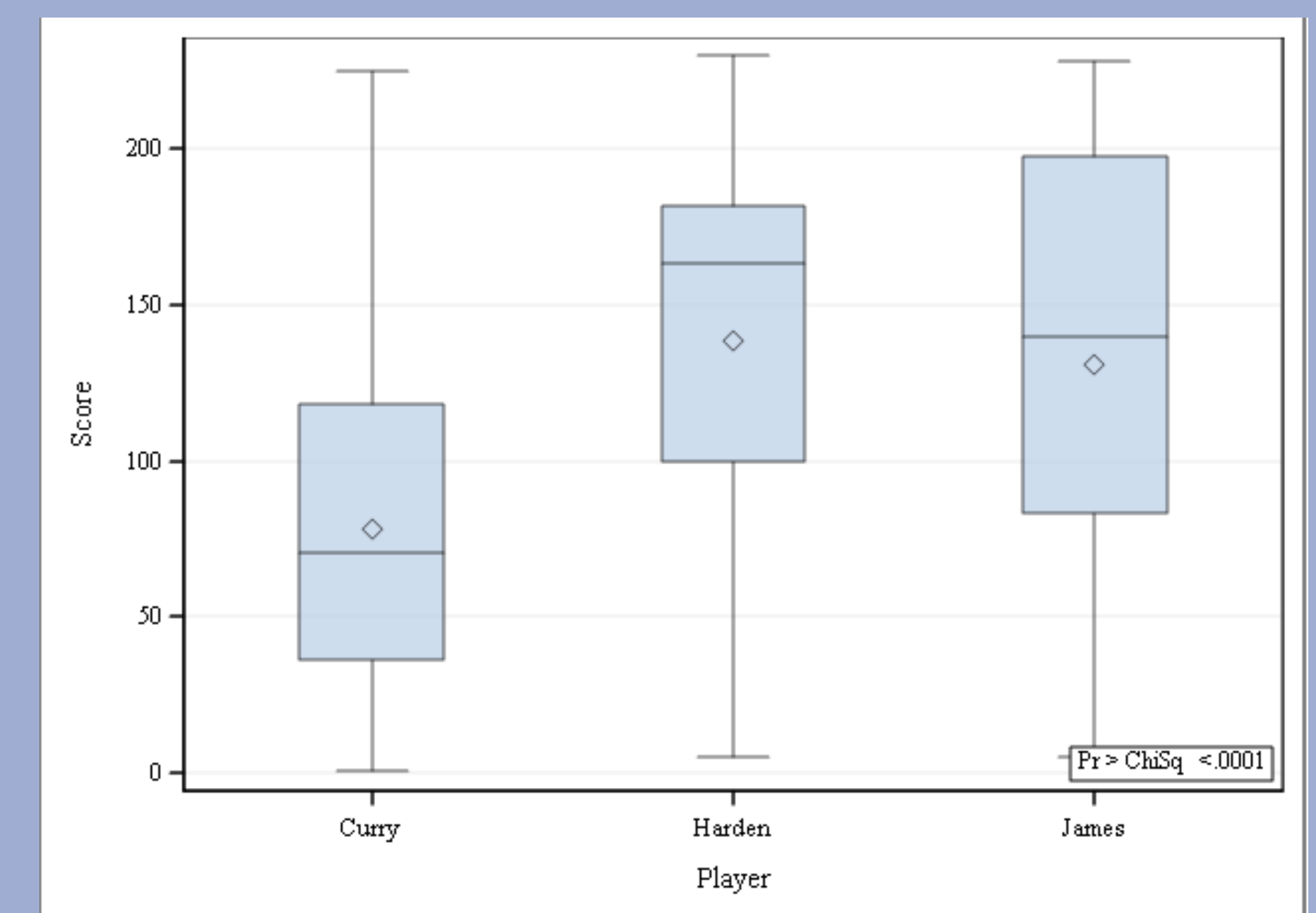
- 2014-2015 season data were retrieved from NBA.com.
- Original dataset: 230 observations and 21 variables
- Removal of redundant explanatory variables
- Standardization of all time-dependent variables
- Retained raw variables for purpose of comparison
- Retrieval of data from past 6 seasons for further analyses

Preliminary results showed a significant difference in the number of minutes played, with Stephen Curry, on average, playing fewer minutes than the other two players.

Table 1. Descriptive Statistics for Minutes Played

Player	N	Minimum	Maximum	Mean	Median	Std Dev
Curry	80	19.00	43.00	32.66	33.00	4.32
Harden	81	25.00	48.00	36.78	38.00	4.10
James	69	25.00	44.00	36.14	37.00	4.68

Figure 1. Boxplots of Minutes Played



Using Parametric and Nonparametric Tests to Assess the Decision of the NBA’s 2014-2015 MVP Award

Sherrie Rodriguez, MS in Applied Statistics

Advising Faculty: Dr. Bradley Barney

Kennesaw State University

II. K-Sample Methods

ANOVA and Kruskal-Wallis tests were conducted on the following variables using an alpha level of .10 for the global tests:

Variable	Description
THREEPERC	Three point percentage
TWOPERC	Two point percentage
FTPERC	Free throw percentage
AST_MIN	Assists per minute
STL_MIN	Steals per minute
OREB_MIN	Offensive rebounds per minute
DREB_MIN	Defensive rebounds per minute
BLK_MIN	Blocks per minute
TOV_MIN	Turnovers per minute
PF_MIN	Personal fouls per minute
TSP	True shooting percentage
PLUSMINUS_MIN	Difference between points scored by his team and points scored by opposing team during time player is on the court

II. Results

- Comparing the p-values of the parametric vs. nonparametric methods, the results were very similar. This demonstrates both the robustness of the ANOVA when the assumptions of normality and equal variances are questionable, as well as the effectiveness of the Kruskal-Wallis test.
- Nine variables were shown to be statistically significant, indicating that there are differences among the performances of the three players. These variables and their p-values are provided in Table 2.

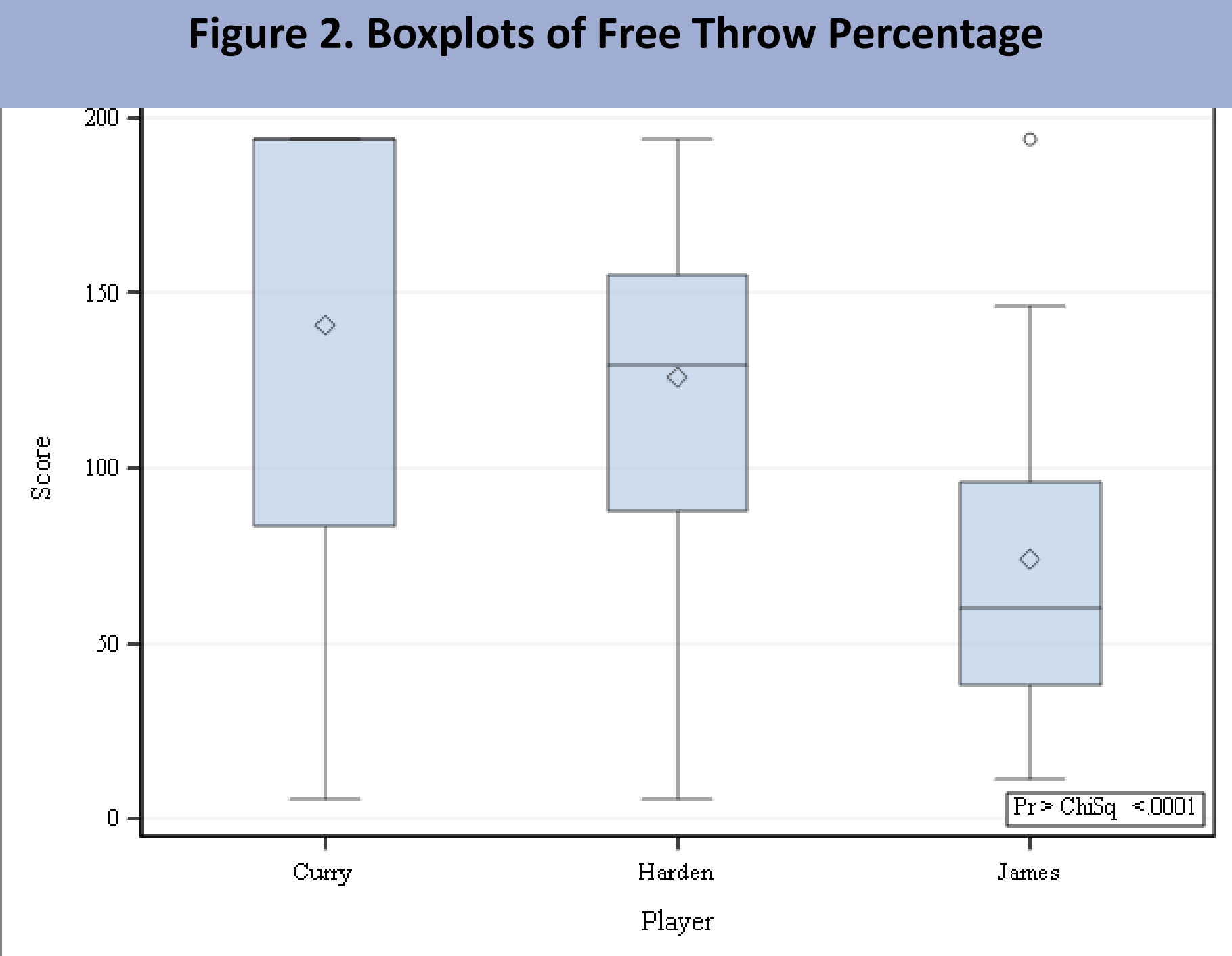


Table 2. Summary of Test Results

Variable	Tests	P-values
THREEPERC	ANOVA	.0113
	Kruskall-Wallis	.0121*
TWOPERC	ANOVA	.0824*
	Kruskall-Wallis	.0327
FTPERC_MIN	ANOVA	.0008
	Kruskall-Wallis	<.0001*
AST_MIN	ANOVA	.0002*
	Kruskall-Wallis	.0010
STL_MIN	ANOVA	.0143
	Kruskall-Wallis	.0111*
DREB_MIN	ANOVA	.0014
	Kruskall-Wallis	.0020*
BLK_MIN	ANOVA	<.0001
	Kruskall-Wallis	<.0001*
PF_MIN	ANOVA	.0338
	Kruskall-Wallis	.0836*
PLUSMINUS_MIN	ANOVA	.0004*
	Kruskall-Wallis	.0006

Using Parametric and Nonparametric Tests to Assess the Decision of the NBA's 2014-2015 MVP Award

Sherrie Rodriguez, MS in Applied Statistics

Advising Faculty: Dr. Bradley Barney

Kennesaw State University

III. Post-Hoc Comparisons

For the variables that satisfied the parametric assumptions, paired comparisons were conducted by using Fisher's LSD through PROC GLM with the option "adjust = t lines". For the distributions that did not meet the assumptions of normality and equal variances, the Wilcoxon Rank-Sum test was the method of choice. The selected alpha level for all paired comparisons was .05.

Discussion

The results of the head-to-head competition between Stephen Curry and James Harden appear to favor Curry. Allow Harden performs better defensively, Stephen Curry comes out on top in four out of six categories using the variables that were found to be statistically significant, three of which relate to offensive skills. The PLUSMINUS variable also favors Curry. The results are similar in the match-up between Stephen Curry and LeBron James. Although James is clearly the better defensive player, Curry beats him in four out of six categories using the variables that were found to be significant. However, the winner for second place is not apparent through the statistical analyses.

III. Results

Table 3. Stephen Curry vs. James Harden

Variable	Test	Best Overall	P-values
THREEPERC	Fisher's LSD	Stephen Curry	.0211
AST_MIN	Fisher's LSD	Stephen Curry	.010
STL_MIN	Wilcoxon with Monte Carlo estimation	Stephen Curry	.0241
DREB_MIN	Wilcoxon with Monte Carlo estimation	Not significant at the alpha level of .05	.0972
BLK_MIN	Wilcoxon with Monte Carlo estimation	James Harden	<.0001
PLUSMINUS_MIN	Fisher's LSD	Stephen Curry	<.0001



Table 4. Stephen Curry vs. LeBron James

Variable	Test	Best Overall	P-values
THREEPERC	Fisher's LSD	Stephen Curry	.0053
FTPERC	Wilcoxon with Monte Carlo estimation	Stephen Curry	<.0001
AST_MIN	Fisher's LSD	Stephen Curry	<.0001
STL_MIN	Wilcoxon with Monte Carlo estimation	Stephen Curry	.0051
DREB_MIN	Wilcoxon with Monte Carlo estimation	LeBron James	.0004
BLK_MIN	Wilcoxon with Monte Carlo estimation	LeBron James	<.0001
PLUSMINUS_MIN	Fisher's LSD	Not quite significant	.0518



Table 5. James Harden vs. LeBron James

Variable	Test	Best Overall	P-values
FTPERC	Wilcoxon with Monte Carlo estimation	James Harden	<.0001
PF_MIN*	Wilcoxon with Monte Carlo estimation	LeBron James	<.0231
TWOPERC	Fisher's LSD	LeBron James	.0333



*Lowest considered "best"

Using Parametric and Nonparametric Tests to Assess the Decision of the NBA’s 2014-2015 MVP Award

Sherrie Rodriguez, MS in Applied Statistics

Advising Faculty: Dr. Bradley Barney

Kennesaw State University

IV. Comparing Standardized vs. Unstandardized Variables

Since there were several time-dependent variables, a comparison was made between the raw variables and a standardized version of each. This was done by dividing by the number of minutes played. The results show two different outcomes between the two versions of the variables. In particular, the outcome between Curry and Harden is significantly different from the previous results, revealing no clear winner. This confirms that the results can be misleading, if the time variable is not taken into consideration.

V. Multivariate Methods

Using NBA data from the past six seasons, PROC CANDISC was used to conduct discriminant analysis. The first discriminant function (Can1) contributes 71.06%, while Can2 contributes the remaining 28.94% (Table 9). These functions were used to plot the player separation shown in Figure 3 using PROC GPLOT. Table 8 displays the standardized coefficients. DREB_MIN (defensive rebounds per minute) contributes the most to Can1. AST_MIN (assists per minute) contributes the most to Can2. Figure 4 shows a contour plot of Can1 and Can2 produced through PROC KDE.

Table 8. Standardized Coefficients

Variable	Can1	Can2
MIN	0.3711786726	0.4105873734
THREEPERC	-.2473685707	0.2497639478
FTPERC	-.1935636695	-.3009517562
TwoPerc	0.2988828358	-.0213128065
AST_MIN	0.0318279811	0.9151733907
STL_MIN	-.1194501520	0.1173345303
OREB_MIN	0.1944370203	-.0679613695
DREB_MIN	0.6511703074	-.1904247356
BLK_MIN	0.2868221796	-.3084199437
TOV_MIN	0.0762985422	0.0800161403
PF_MIN	-.4040826199	0.0676386252
PLUSMINUS_MIN	0.0925445722	-.2096695105

Table 9. Canonical Discriminant Functions

Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation	Eigenvalues of Inv(E)*H = CanRsqr/(1-CanRsqr)			
				Eigenvalue	Difference	Proportion	Cumulative
0.596746	0.591344	0.017858	0.356106	0.5531	0.3278	0.7106	0.7106
0.428736	0.422554	0.022637	0.183814	0.2252		0.2894	1.0000

IV. Results

Table 6. Stephen Curry vs. James Harden (Standardized)

Variable	Test	Best Overall	P-values
THREEPERC	Fisher’s LSD	Stephen Curry	.0211
AST_MIN	Fisher’s LSD	Stephen Curry	.010
STL_MIN	Wilcoxon with Monte Carlo estimation	Stephen Curry	.0241
DREB_MIN	Wilcoxon with Monte Carlo estimation	Not significant at the alpha level of .05	.0972
BLK_MIN	Wilcoxon with Monte Carlo estimation	James Harden	<.0001
PLUSMINUS_MIN	Fisher’s LSD	Stephen Curry	<.0001

Table 7. Stephen Curry vs. James Harden (Unstandardized)

Variable	Test	Best Overall	P-values
THREEPERC	Fisher’s LSD	Stephen Curry	.0211
AST	Global test	Not significant	.1925
STL	Wilcoxon with Monte Carlo estimation	Not significant	.5560
DREB	Wilcoxon with Monte Carlo estimation	Harden	.0024
BLK	Wilcoxon with Monte Carlo estimation	James Harden	<.0001
PLUSMINUS	Fisher’s LSD	Stephen Curry	<.0005

V. Results

Figure 3. Plot of Can1 vs. Can2 to Show Player Separation

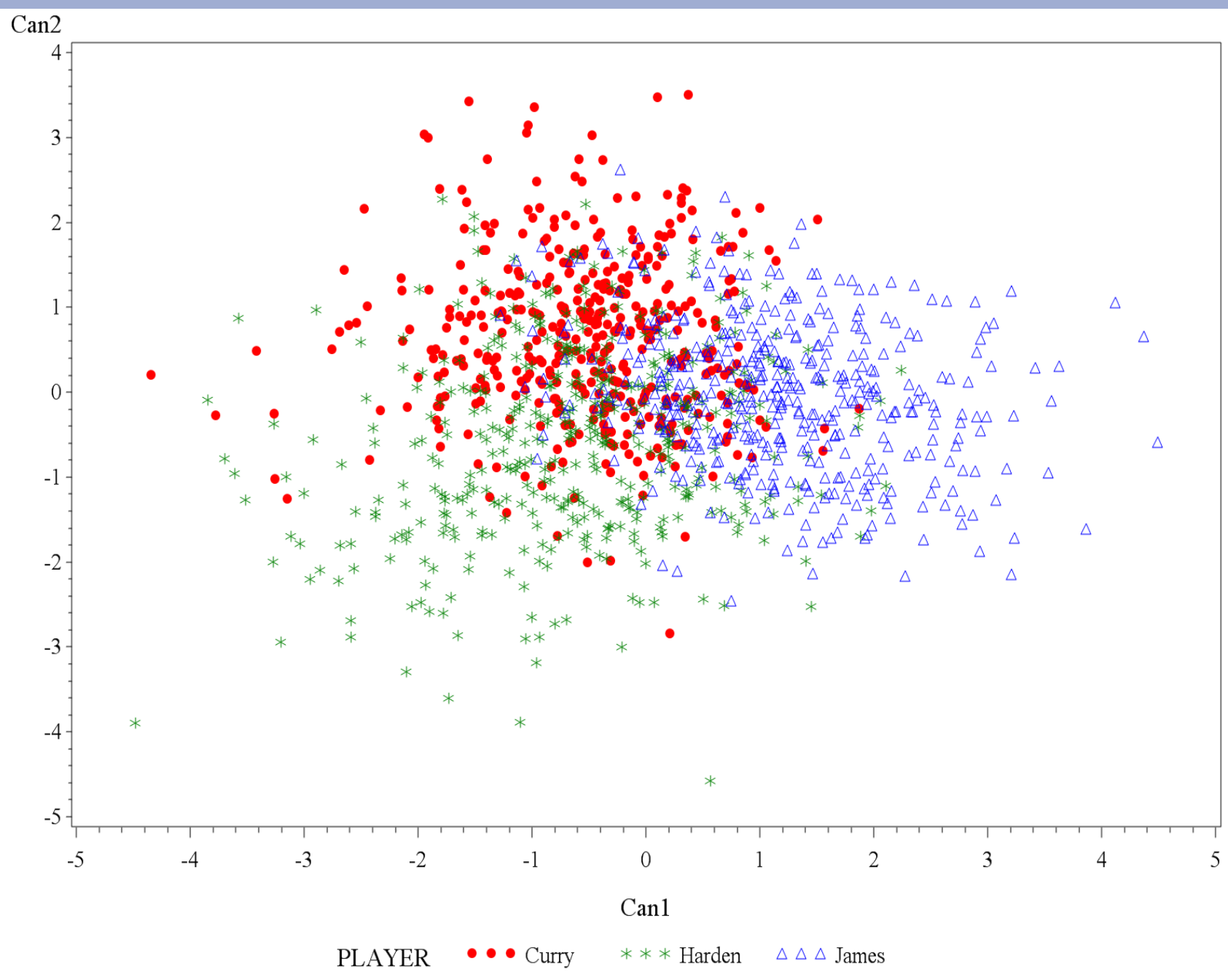
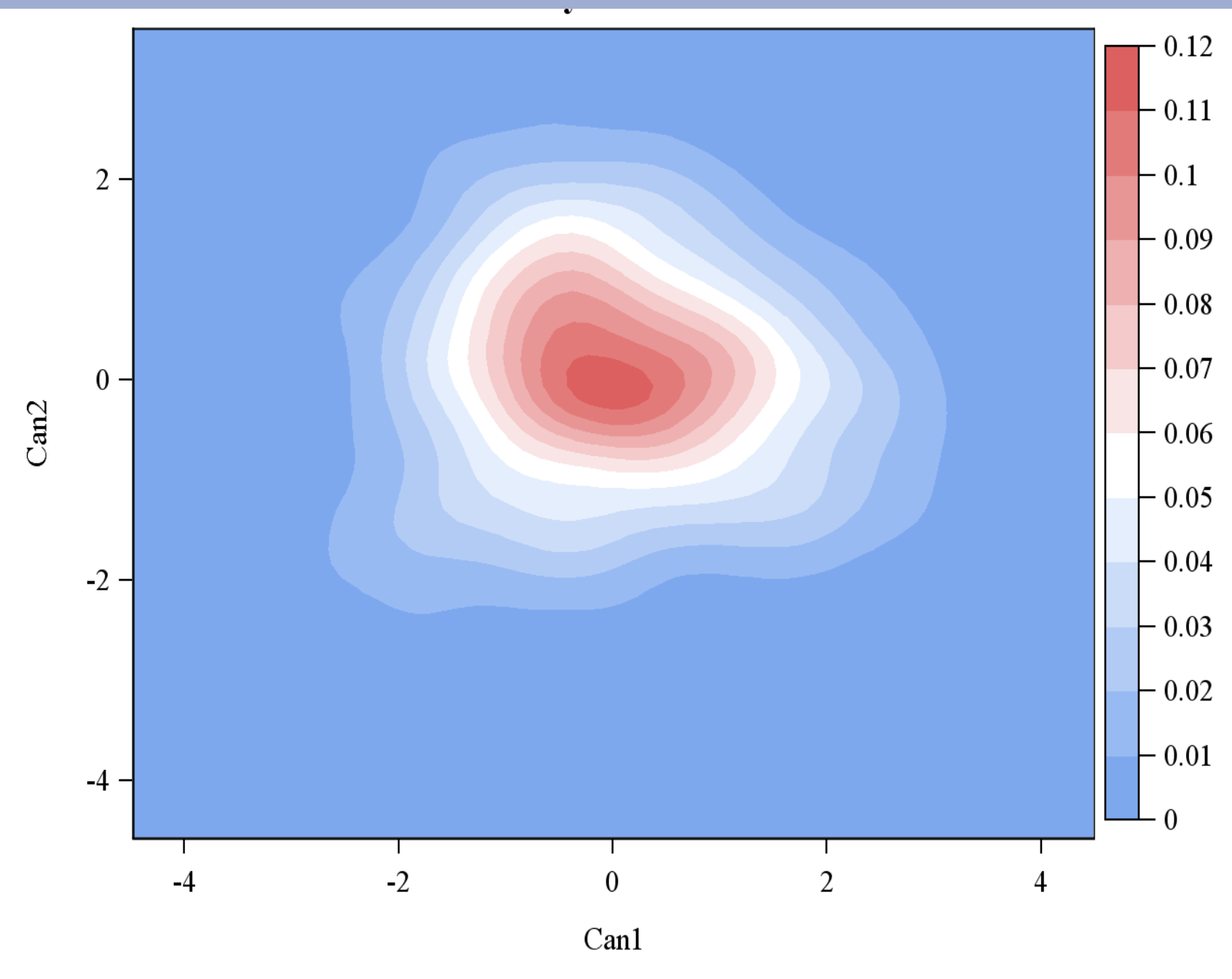


Figure 4. Contour Plot of Can1 vs. Can2

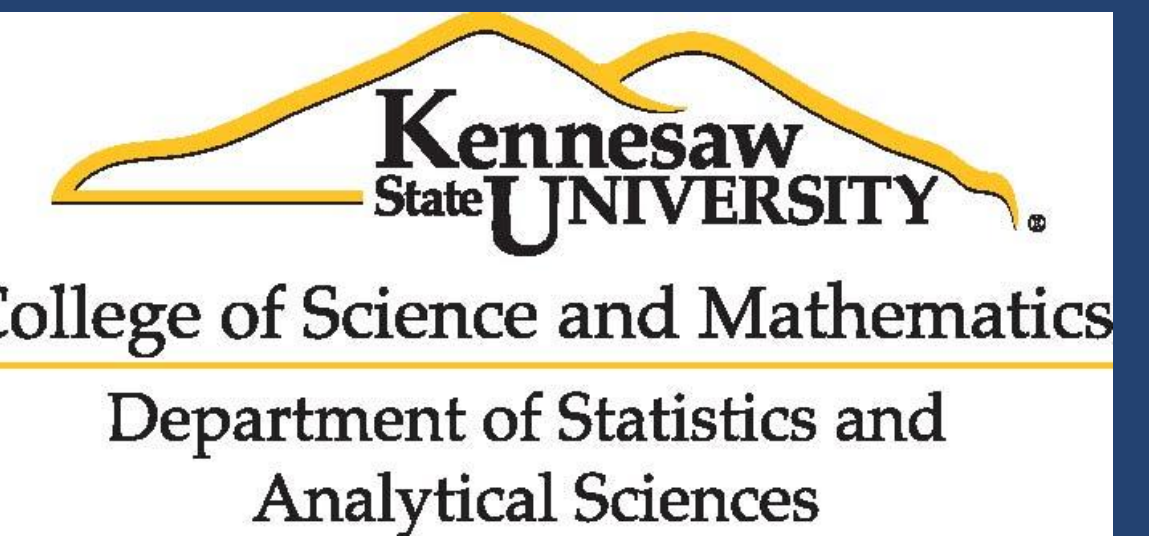


Using Parametric and Nonparametric Tests to Assess the Decision of the NBA's 2014-2015 MVP Award

Sherrie Rodriguez, MS in Applied Statistics

Advising Faculty: Dr. Bradley Barney

Kennesaw State University



Conclusions

Preliminary test results indicated that there was a significant difference in the number of minutes played by each player. In order to account for this, the time-dependent variables were standardized by dividing by the number of minutes.

The results of the global tests indicate that there are significant differences in the players' performances in the following skills: three point percentage, two point percentage, free throw percentage, assists, steals, defensive rebounds, and blocking. Additionally, there is a statistically significant difference with regards to the plus-minus variable. Although the paired comparisons between James Harden and LeBron James resulted in a statistical tie, the post-hoc comparisons between Stephen Curry and the other two players clearly indicate that Curry out-performed both Harden and James. Thus, the statistical evidence confirms the decision of the panel of judges: The 2014-2015 MVP award rightly goes to Stephen Curry.

Comparing the raw time-dependent variables with the standardized version of each, the analyses indicate that there is a significant difference in outcomes if the time variable is not taken into consideration. Had these variables not been standardized, the analyses would have resulted in a statistical tie between Stephen Curry and James Harden.

The results of the multivariate analyses indicate that there is not a great deal of separation between the three players. This is not surprising, given the fact that these were the top three contenders in the NBA during the 2014-2015 season. The 2015-2016 season is likely to be an exciting one, with such talented players in the league.

Acknowledgements

I would like to acknowledge all of the faculty of Kennesaw State University's Department of Statistics and Analytical Sciences who have contributed to my knowledge of the statistical methods used in these analyses. Through the M.S. in Applied Statistics (MSAS) program, I have learned a great deal that will help equip me in my pursuits in the future. I would especially like to thank Dr. Brad Barney, my advising professor for this project.



References

Higgins J. 2004. Introduction to Modern Nonparametric Statistics. California (CA): Brooks/Cole-Thomson Learning. Chapter 3, K-Sample Methods; p. 86-90.

Rencher AC, Christensen WF. 2012. Methods of Multivariate Analysis. New Jersey (NJ): John Wiley & Sons, Inc. Chapter 8, Discriminant Analysis: Description of Group Separation; p. 298.

National Basketball Association: 2009-2015 Regular Season Game Logs. www.nba.com.

NBA News Kia MVP Official Release: May 4, 2015. www.nba.com/2015/news/05/04/stephen-curry-2014-15-kia-mvp-official-release

http://www.bet.com/news/sports/photos/2015/04/2015-nba-playoff-preview/jcr_content/leftcol/flipbook/flipbookimage.flipfeature.dimg/041715- Web. 7 Sept. 2015.

http://img.bleacherreport.net/img/images/photos/003/240/147/9bc8daeaf123c4db48c258ea4294f504_crop_north.jpg?w=630&h=420&q=75 Web. 12 Sept. 2015.

http://assets.rappler.com/612F469A6EA84F6BAE882D2B94A4B421/img/62BF09C757D34A1F99F9B72D4B549B9A/james-curry-20150531_62BF09C757D34A1F99F9B72D4B549B9A.jpg Web. 12 Sept. 2015.

http://cdn.abclocal.go.com/content/ktrk/images/cms/540170_1280x720.jpg Web. 12 Sept. 2015.

http://i.cdn.turner.com/drp/nba/warriors/sites/default/files/styles/story_main_photo/public/kia-mvp-curry-760x442.jpg?itok=RmsmlNj2 Web. 12 Sept. 2015.



SAS[®] GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.

LAS VEGAS | APRIL 18-21

#SASGF