Predict Warriors' 2015-2016 73 Win Record

Mason Chen
Milpitas Christian School, Morrill Learning Center
San Jose, CA 95132
mason.chen.training@gmail.com

Zonghuan (Jason) Li
San Mateo High School, Morrill Learning Center
San Mateo, CA 94401
jasonlzh2001@gmail.com

Abstract

Build a predictive model to simulate the probability whether 2015-2016 Warriors can break Chicago Bull's 1995-1996 72-Win record on April 13, 2016. Team used the Yahoo NBA live database when Warriors have completed their first 50 games on Feb.6, 2016 to build a predictive model to simulate winning 73 games. Sample size 50 was determined through Power Test to derive the 1-sided Proportion Lower Confidence Interval in 70% to 99% confidence level range. Proportion Confidence level was calculated with Approximately Normal Z Statistic. Team monitored the 73W probability after Feb.6, and decided when to purchase April 13 82nd game tickets through an analytical decision tree. The control chart analysis has explained Warriors' tough adventure when approaching the 73W history on April 13. Team was able to witness and celebrate this Warriors 73W History at Oracle Arena. The predictive model may not be very accurate all the times, but the ending was surprisingly matching the prediction made in early Feb. There are always certain degrees of risk associated with random variations. Team has made two smart decisions: (1) predicted Warriors to win 73 Games, and (2) purchased tickets earlier.

Keywords
Modeling, Statistics, Decision Tree, Control Chart, Minitab

1. Project Introduction
With the increased prominence of such large data sets, it was inevitable that state-of-the-art machine-learning techniques would begin to make their mark in Sports Analytics. Team members are Warriors' fans living in the highly-stressed Bay Area. Parents are working so hard in order to earn life in Bay Area. Warriors' amazing performance this year has happily changed life dramatically. After unbelievable 23-0 opening record, people here are talking about "73W" on April 13, 2016. In late January, in one Birthday Party, team discussed how to predict the 73W probability in Statistics and Probability. After the party, team has formed an IEOM STEM Project Team in early Feb. to study and research statistics on simulating the Warriors' winning probability of each remaining game.

1.1 Project Hypotheses
After team brainstorming, team has identified two major hypotheses:
1. Can team use basic probability to predict whether Warriors can win 73 Games on April 13, 2016?
2. When should team purchase tickets of the final Warriors' 82nd regular game?

1.2 Project Literature Research
Before team would build the predictive model, team has searched the NBA website and found a similar approach of predicting Warriors' 73Wins probability at 538 Prediction [1]. See Figure 1, the 538 prediction has indicated that Warriors could win 73 games, more exactly at 73.5 games in the season end. Team did not know how this model coming out and team would use this model as a baseline reference when building the predictive model. This 538 model did not include the confidence interval concept and team could certainly build a better model with adding the confidence interval statistics. A research has studied what separates a player evaluation scheme from previous ones

on identifying the substitutability between different game actions, and then using that to quantify the complementary effect teammates have on one another [2,3,4].
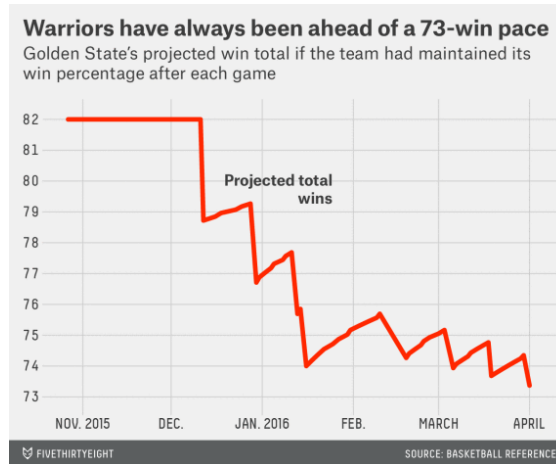

Figure 1. 538 Prediction Model

1.3 Build Warriors Project Team

Project mentors suggested that I could form a Statistics Team to compete the IEOM STEM contest. Team has formed a STEM Project Team of one leader and two mentors. Below is the role and responsibility of each team member:

- Leader: collect Warriors' data and information including Warriors' schedule and game result. Use AP Statistics and Minitab software to build a predictive model and update Statistical analysis when needed.
- Mentor A: provide statistics guidance and training if needed
- Mentor B: Warriors Fan for 40 years. Provide Warriors' players information and introduce Coach Kerr.

2 Build Model

Team then decided to use Warriors' first 50 games' record and other teams' record available on Feb.6 to establish the predictive model. Team collected the first data on Feb.6 after Warriors have just completed their 50th Game. Team also updated the predictive model after completed each Warriors' remaining game in order for team to decide when team should purchase tickets of Warriors' final 82nd Game. Team doesn't want to miss this probable Once-In-Your-Life opportunity. If team would purchase tickets too early, the Warriors may lose more games and never reach 73win on April 13. If team would purchase tickets too late, the Warriors may be likely to win 73 games and the ticket price may rock up. Therefore, team also monitors the 82nd Game ticket price after every game.

2.1 Data Collection

When Warrior completed the first 50th Game on Feb.6, team could collect each NBA team's record (including home game record and guest game record). For example, Warriors have 46 Win-4 Loss Record (home record is 23 Win-0 Loss and guest record is 23 Win-4 Loss). Team discussed the sample size and thought 50 games should be adequate to simulate the Binomial Confidence Interval then (see the analysis on Page 5). Team built the first predictive model on Feb. 7 based on teams' accumulated record in the end of Feb.6 available at Yahoo NBA Standing Record Page [5].

2.2 Model Algorithm

Team calculated the Warriors' winning probability of each remaining game by the following formula:

- If game is at Warriors' home court, team used Warriors' home record and the opponents' guest record to determine Warriors' winning probability
- If game is away from Warriors' home court, team used Warriors' guest record and the opponents' home record to determine Warriors' winning probability

- Team also used the sample size and accumulated winning % to simulate the 1-sided Lower Confidence Interval of Warriors' winning probability of each game

After some brainstorming sessions and consulted with two mentors on Statistics, team has come out with a predictive model as following: team will use the Table 1 to demonstrate the Predictive Model.

Table 1. Model Algorithm

A	B	C	D	E	F	G	H	I	J	K
Home Game Opponent	Guest Game Opponent	Warriors' Record			Opponent Record			Warriors' Win%	Warriors' Lose%	Warriors' Winning Probability
New Orleans		23	0	98%	6	21	23%	75%	0%	0.99
New York		23	0	98%	9	17	34%	64%	1%	0.99
	Dallas	23	4	85%	17	10	61%	33%	9%	0.79
	San Antonio	23	4	85%	26	0	98%	1%	14%	0.09

Team listed the accumulated record of Warrior and Opponents' home and guest game record and information on Feb.6 (columns A, B, C, D, F, and G). Then, team calculated the Warrior's winning probability based on the following steps:

- In column E, team calculated Warriors' current Winning % based on the home record or guest record of each remaining game accordingly. Since Warrior had a perfect home record on Feb.6, team was assuming Warrior has lost 0.5 game (to avoid singularity in calculation) to assign 98% winning% in Table 1.
- In column H, team calculated Opponents' current Winning % based on the home record or guest record of their particular game playing with Warriors accordingly. Since San Antonio Spurs also have a perfect home record on Feb.6, team was assuming Spurs have lost 0.5 games then to assign 98% winning%.
- In column I, team calculated Warriors' Win% under the condition if Warrior Win AND Opponent Lose. The Column I = [Column E] * [1- Column H]. Team used conditional probability and assumed both events were independent. Warriors have played 50 games and have played with the other teams once or twice before Feb.7. The dependency degree is below 5% (2/50= 4%). Two mentors agreed that team can assume this independence at a very low risk in the predictive model.
- In column J, team calculated Warriors' Lose% under the condition if Warrior Lose AND Opponent Win. The Column J= [1- Column E] * [Column H]
- In column K, team determined Warrior's Winning Probability for any particular remaining game as Column K= [Column I + Column J]/[Column I]

Based on the above predictive model algorithms, team can calculate the winning probability of each Warriors' remaining 32 games after Feb.6. After some preliminary calculation, team would predict that Warrior can win **74.08** games on April 13, 2016. The 74.08 game prediction was close to the 538 professional prediction at 73.5 games in Figure 1. Wow, team could be professional on predicting the Warriors'73W. The ESPN should invite us to present model in each Warriors' game. Though, team was not satisfied with current point estimation result. Team wanted, further, to use the inferential confidence interval algorithm to enhance prediction capability.

2.3 Build the Confidence Interval

Team was very excited by this proto model analysis. Next question, what's the confidence of Warriors' winning at least 73 games on April 13, 2016. Previous 74.08 games prediction is assuming there is no random variation and confidence at 50% level. From AP Statistics, team has learned Inferential Statistics in addition to the Descriptive Statistics. Team was curious how to apply the Confidence Interval into the predictive model. Team consulted with Mentors and asked what Inferential Statistics can be used to determine the confidence interval.

After consulting with Mentors, team has decided to use the following statistics to determine the confidence level of winning of 73 games on April 13.

- Use **Normal Z Approximation** to simulate the **Binomial Distribution** [6]: sample size is > 50 and most winning % records are between 20%-80%. Team checked two "Normal Approximation" requirements (1) np= 50 *0.8 = 40 ≥ 10, and (2) nq= 50 *0.2 ≥ 10 (to avoid asymmetry concerns of Normal Approximation).
- The above analysis has supported earlier decision that 50 game sample size was good sufficient for Normal Approximation.
- Team used combined winning percent Pbar of each game from both Warriors and Opponent to simulate the 1-sided confidence interval (since team only cared the lower limit whether Warrior can win at least 73 games).
- Team used Normal Approximation to derive the Binomial Standard Deviation=

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

- Then, team calculated the 1-sided Confidence Interval as following: confidence= 1- Alpha Risk. Team used Z table to determine the Z_α Statistic in order to calculate the 1-Proportion Z Confidence Interval [7].

The lower Confidence Interval limit of Winning Probability =

$$\text{Lower Confidence Interval of } \mu = p - Z_\alpha \sqrt{\frac{p(1-p)}{n}}$$

- Then, team accumulated each Lower Confidence Interval Limit of each remaining game to determine the overall Lower Confidence Interval Limit of the eventual 82nd Game Interval.
- Based on Feb. 6 record and the above confidence interval algorithm, team has come out with the following Confidence prediction of winning 73 games in Table 2.

Table 2. Confidence Interval Model of 73Win Prediction

Lower Confidence Interval of Winning Games on April 13						
70%	75%	80%	85%	90%	95%	99%
73.63	73.55	73.46	73.36	73.23	73.04	72.68

In Table 2, the predictive model has indicated that Warriors have more than 95% to win 73 games this season, but not at 99% confidence yet. Team was so excited that in about two months, Warriors are likely going to make a history. Team is so lucky that they are living in the Bay Area and able to witness this history. Team was also very happy that team could apply AP Statistics Learning in this project. Mentor has taught us AP Statistics in a more practical way in a project-based learning process. Team has created an Excel file and it would only take team about 10mins to update the predictive model after Warriors have completed each remaining game. Team would update Warriors' record and each opponent most updated record to recalculate the winning probability of each remaining game.

3. Develop Decision Tree of Purchasing Warriors' 82nd Game Tickets

The next challenge was to determine when team should purchase tickets of Warriors' 82nd Game on April 13. Team checked the ticket price, in middle Feb., on the NBA. Net and the cheaper tickets in the higher Sections were around $200, and the Lower Sections were all above $500. Team knew Warriors just won the Champion last year after 40 years. Team also knew that every NBA fan has been waiting for 20 years to see whether Warrior can break Michael Jordan's Bull 72W record. Team also knew that tickets may go much higher if Warriors can keep winning. Should team take a higher risk to purchase tickets now to get a possible higher return on April 13? Or, should team wait a little bit longer, and collect more sample size to ensure our bet on Warriors' 73W more reliably. Uh, it's a dilemma.

3.1 Risk Assessment

Team consulted with mentors again, and they suggested that team can conduct a risk assessment and identify the most risky remaining games and schedule. Based on their suggestions and predictive model, team came out with the seven most risky games in the following list:

1. March 3 Game #60 vs. Thunders at home.   Winning Probability = 87%.
2. March 18 Game #68 at Dallas.   Winning Probability= 79%
3. **March 19 Game #69 at Spurs.**   **Winning Probability= 9%.**
4. March 30 Game #75 at Utah.   Winning Probability= 77%.
5. April 7 Game #79 vs. Spurs at home.   Winning Probability= 85%.
6. April 9 Game #80 at Memphis.   Winning probability= 70%.
7. **April 10 Game #81 at Spurs.**   **Winning Probability= 9%.**

When team looked at the list, what, Warriors still needed to play with Spurs three times and two at their home court. Warriors had only 9% to win at Spurs' home court. Based on risk assessment, with up to 5 more losses to give (Warriors could lose total 9 games and Warriors has lost four games already), Warriors may allow one unexpected loss to the other weaker teams not in the above 7 games (this probability is determined based on Warriors' first 50 games, Warriors could not win every easy game). Warriors would possibly lose 2 of 3 games playing with Spurs (#69, #79, and #81). Then, Warriors could afford two more losses among the following four critical games:

1. March 3 Game #60 vs. Thunders at home.   Winning Probability = 87%.
2. March 18 Game #68 at Dallas.   Winning Probability= 79%
3. March 19 Game #69 at Spurs.   Winning Probability= 9%.
4. March 30 Game #75 at Utah.   Winning Probability= 77%.
5. April 7 Game #79 vs. Spurs at home.   Winning Probability= 85%.
6. April 9 Game #80 at Memphis.   Winning probability= 70%.
7. April 10 Game #81 at Spurs.   Winning Probability= 9%.

3.2 Set Decision Points

Based on the schedule, team has decided to set the following three decision points:
(1) If Warriors could win 9 in next 10 games (#51- #60), team will purchase 82nd game tickets on March 3 after Warrior could win over Thunder at Game #60.
(2) If Warrior unfortunately lost more than one game before March 4, team would wait until next check point after completed two critical games: March 18 Game #68 at Dallas and March 19 Game #69 at Spurs. If the 73W confidence was above 95% level, team would purchase tickets after Game #69 on March 19.
(3) If team might still fail to make a decision on March 19, team would conduct another risk assessment on March 30 after Utah Game to determine whether and when to buy tickets, and so on…

3.3 Create Decision Flow Chart

Tam has created a Flow Chart Diagram to demonstrate three Decision Points in Figure 2. If team could not find decision on March 30, team would schedule a special meeting on March 31 to discuss whether team could take a risk to make a deviation from the predictive model and flow chart, and purchase tickets on March 31. Would team be willing to take a higher Alpha Risk (lower Confidence level)? At that moment, team was really hoping that Warriors could keep winning and team won't face this critical situation to lower the confidence level. At the same time, team wanted to purchase tickets earlier in order to get a good deal.
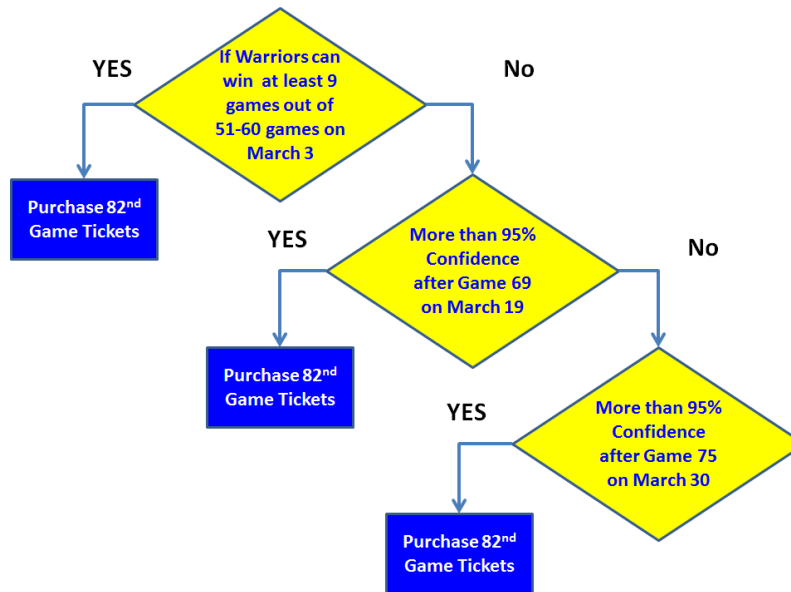
Figure 2. Decision Flow Chart

4. Verify the Predictive Model

After completed the above analysis, team has updated the 73-Game Lower Confidence Limit after completed each Warriors' Game. The confidence level of winning 73 games was monitored frequently, and discussed among team members and mentors. Life was totally changed from Feb. 6. In school, team would share the predictive model and result with classmates. At home, team would watch each Warriors' game with parents. Team has created a lot of friendship through this special Warriors' 73W model.

4.1 First Loss to Portland on Feb.9

Warriors kept wining until lost one game at Portland on Feb. 19. Team was excited to see how much impact of this loss based on the predictive model. Fortunately, team still had 95% confidence that Warriors can win 73 games in the end (see Table 3 below). Team was not panicking anymore since the model would tell team how much impact of winning or losing each game. Team was more excited if Warriors would play with a tougher team and Warriors have a better chance to boost their confidence level. In the other side, team also worried that Warriors might lose to a weaker team and dropped their confidence level significantly. When team watched each Warriors' game, team used model to assess the impact of each game result. Warriors were able to pull several games in the overtime. Could anyone imagine what our mind was thinking at that time? Please don't lose tonight, and this loss may impact the confidence by 5-10%. Following this predictive statistics, watching each Warriors' game was becoming so enjoying and engaged on each move on Court. Come on No. 30 Curry!

Table 3. 73Win Prediction on Feb.19

Lower Confidence Interval of Winning Games on April 13				
70%	80%	90%	95%	99%
73.45	73.36	73.28	73.06	72.52

4.2 First Decision Point on March 3

On March 3, a big night (first check point in Decision Tree), Warriors will play with mighty Thunders (3rd Best Team in the West). If Warriors could win over Thunders, team would purchase 82nd game tickets immediately no matter what the ticket price would be (team would save spending on Video Games or other Shopping). By the way, team members had been reducing Video Game playing significantly. Team has been spending more time on watching Warriors' game, updating prediction, and studying Statistics. The relationships with parents have been much better aligning with this common goal: (going to see Warriors' 82nd game on April 13 to break Jordan's

record). If Warriors could win Thunders on March 3, based on the most updated model (see Table 4 below), Warriors has "99%" confidence to win 73 games. The chance on March 3 was even higher than Feb.7 result.

Table 4. First Decision Point on March 3

Lower Confidence Interval of Winning Games on April 13				
70%	80%	90%	95%	99%
74.00	73.83	73.59	73.39	73.02

As expected, Warrior had a big win over mighty Thunders (121-106). Team immediately purchased tickets ($231/ticket). Team had been waiting for this Big Game and team were so excited and so happy to spend such Big Money on Warriors. They deserved each penny. When team confirmed the ticket purchase, something came out from mind: are we crazy? Does team really believe in Statistics? Team shared this crazy story with classmates and friends next morning.

4.3 Enjoy Statistical Uncertain Random Variation

Yeah, they might be right. Warriors lost to the worst team Lakers in the West on Game #61 March 6, just three days after team purchased tickets. Was that a Joke? Team could not hide in school and every close friend was laughing at us. Team heard a lot of bad comments on Curry and Warrior Team about their weakness from the NBA Experts…. Warriors' season schedule was too easy so far. Their real tough test was coming. Even Pippen and Barkley (formal NBA stars) claimed that Jordan's 72W could sweep today's Warriors' team 4-0 in the Playoff. Suddenly, team morale was so low at that time. Team mentors encouraged us: you got to believe in Statistics and more importantly, believe in Warriors. Today's NBA playing style is totally different from Jordan's era in 20 years ago. Warriors were creating their own winning style and no true expert could judge what might happen now on Warriors. They would find a way to win 73 Games on April 13 to become one of the greatest teams in NBA History. Team went back and updated Model (Table 5) after this terrible lost to Lakers. Mentors were right. Even lost to Lakers at Game #61, they still had 95% confidence to win 73 games in the end. Team got to believe in Warriors. They were not happy about this loss. Coach Kerr will find a way to overcome this loss.

Table 5. 73Win Prediction on March 6

Lower Confidence Interval of Winning Games on April 13				
70%	80%	90%	95%	99%
73.86	73.64	73.44	73.24	72.87

4.4 April 1st Fools Day

In the next month, team was experiencing in an up and down **Roller Coaster mode**. Warriors won some tough games but also lost to some much weaker Opponents as unexpected. On April 1 (Fools Day), Warriors was basically fooled by Boston Celtics' team. Boston was a good team but no one would expect they could beat Warriors at home. Before this game, Warriors had a perfect home record and had won 53 home games consecutively (also NBA record). Team was anxious to see how much impact of this loss based on the model. Suddenly, this was the first time that the predictive model had shown that Warriors would have less than 95% confidence (see Table 6 below) to win 73 games. **Team only had 80% confidence at this moment.** Typically, team wanted to make a critical decision at 95% confidence. Team knew today was "Fools Day". Everything should be OK tomorrow when team wakes up, right!

Table 6. 73W Prediction on April 1

Lower Confidence Interval of Winning Games on April 13				
70%	80%	90%	95%	99%
73.44	73.26	72.98	72.92	72.48

Four days later, on April 5, Warriors lost #78 game to Minnesota. This time, the confidence level was dropped to below 70% (see Table 7 below, maybe in 50-60% range). Today was not the "Fools Day". 50% confidence is basically like playing Las Vegas gambling. Warriors had no excuse why Warriors lost again to a very weak team (lost two home games in five days). Coach Kerr said this was natural since players were nervous now. He went through this experience 20 years ago when Bulls built 72-W record. Come on, Coach Kerr. Please don't find any excuse. Please do something now! Should team still believe in Coach Kerr and Warriors?

Table 7. 73W Prediction on April 5

Lower Confidence Interval of Winning Games on April 13				
70%	80%	90%	95%	99%
72.85	72.66	72.37	72.28	71.89

Warriors' record was 69-9 now, and Warrior needs to win the ALL remaining four games including two games vs. Spurs and one game vs. Memphis. April 13 game was just about one week away.

April 7 Game #79 vs. Spurs at home.   Winning Probability= 85%.
April 9 Game #80 at Memphis.   Winning probability= 70%.
April 10 Game #81 at Spurs.   **Winning Probability= 9%.**

4.5 Warriors' 73 Win

Warriors could not miss any game in these three BIG games. Warriors were able to pull each game one-by-one and won all three games in four days. During that span, team had a tough annual school exams week before the Spring Break. Team cared more on the Warriors record than individual school exams. After won three Big Games on April 10, based on the most updated prediction (Table 8), team has 95% confidence that Warriors can win the last game and set the new record on April 13. Warriors would fly back to Bay Area and had three full days to rest before the Final 82nd Game.

Table 8. 73W Prediction on April 12

Lower Confidence Interval of Winning Games on April 13				
70%	80%	90%	95%	99%
73.67	73.48	73.19	73.05	72.77

BIG DAY was finally coming on April 13. Team was so excited and could not wait to see the Warriors' Oracle Arena. The ticket price was rocking up every hour since April 10. Team could sell each ticket at least $500 but team won't sell any since this ticket was how team believed in Warriors and Statistics. Team had a big plan to enjoy every minute on. Every seat was taken. Every fan was crazy there. Game was so special, and so many wonderful Game 82. The traffic on HW 880 was so bad in the late afternoon since so many cars were driving to Oakland. Team just arrived 10 minutes before the Game Start. Each of us got the 82-Game T-Shirt on the Chair. Team got a free Clips haircut coupon in the end. Curry was crazy (dropping almost every 3-point). Game was basically determined after 20-minutes. Everyone stayed late and team wanted to celebrate this history with Warriors Superstars. They deserved each screaming in this game. Team members lost voices and got exhausted during the whole game.

5 Post-Game Model Validation Analysis

After completed Game 82, team was eager to know whether the predictive model could explain what happened on Warriors' Road to 73Win.

5.1 Scatterplot Analysis

Team had been monitoring the confidence after completed each game. In Figure 3, the scatterplot had shown the projected winning games vs. Warriors' completed games at different confidence interval. The red thick dot-line was 73 games as Reference. The thicker purple marker was at 95% confidence.
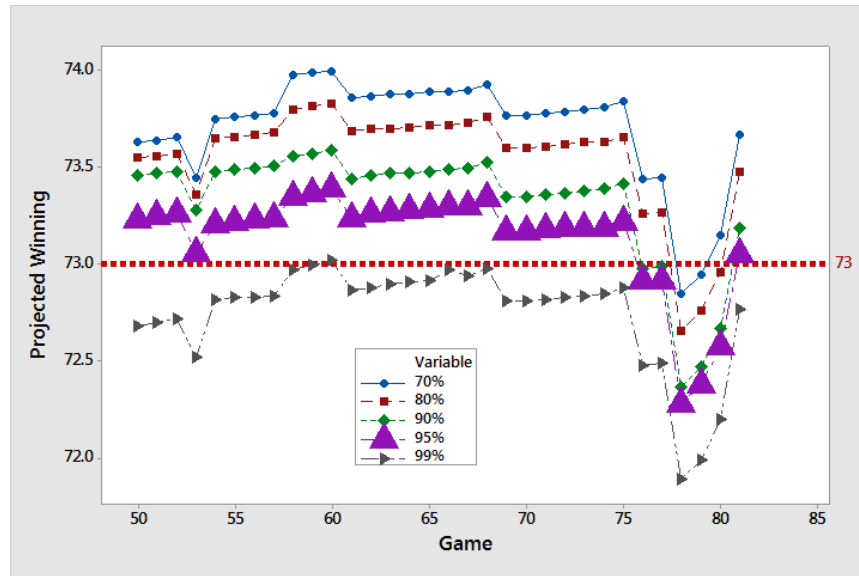
Figure 3. Scatterplot of Prediction

From the plot, Warriors had being projected to win 73 games at 95% confidence most time before the final 6 games. After lost to Boston and Minnesota in early April, the impact was significant. Though, the Warrior could put a team together quickly and won the next three tough games. The curve has shown all this V-shape drama in the last seven games. Warriors have a very stable performance in the entire season except the final 7 games, why? Like Coach Kerr said: they were experiencing something never happened in their professional life. They cared something which can make history. They were nervous on court, and they needed to find a way to make a history if they really wanted to do it as a team together. Curry changed his style and assisted more players to score. They protected their defense rebounds and reduced their passing errors. They tried to control their tempo by better communication to avoid being too excited or too nervous. Mentor made such comments. They eventually made it and the curve had shown how they can resolve or eliminate this abnormal (special variation) quickly. Only a great team could do it quickly. It's their minds never giving up and believing in themselves. It's us (Warriors' fan) who also believed they would make a history on April 13.

5.2 Control Chart Analysis

Mentor also suggested us that team could use a I-MR Control Chart [5] concept to examine this unusual behavior shown in Figure 4. Team got trained on using Minitab software from Mentors and conducted the following Control Chart Analysis. There were two control charts team used: (1) Individual Chart which plotted the projected wins at 95% confidence level, and (2) moving range chart which plotted the moving range between two consecutive games. The individual chart would indicate chance of winning 73 games at 95% confidence. Both charts would use the upper control limits and lower control limits to identify the outliers. The moving range chart would indicate whether Warriors were performing stably if not winning or losing any unexpected game based on the predictive model. Based on Figure 4, team could observe several out-of-control (OOC) points. OOC points were like outliers. Team was curious how these OOC points would tell what story. The first OOC point on the Individual Plot is at Game 60. This point was outside the upper control limit, which means an outlier above the upper control limit. Mentors told us this indicated the probability of being outside control limits was very rare below 1%. Interestingly, team set the first check point at Game 60 when Warrior would play with Dallas on March 3rd. Lucky or unlucky, first check point for purchasing tickets was on the unusual higher side, which means over-confident to make a critical decision. If based on this control chart, team might not make a decision on Game 60, and may wait another one or two games when the curve would go down below the upper control limit. Game 61 might be a better choice for making the ticket purchasing decision.
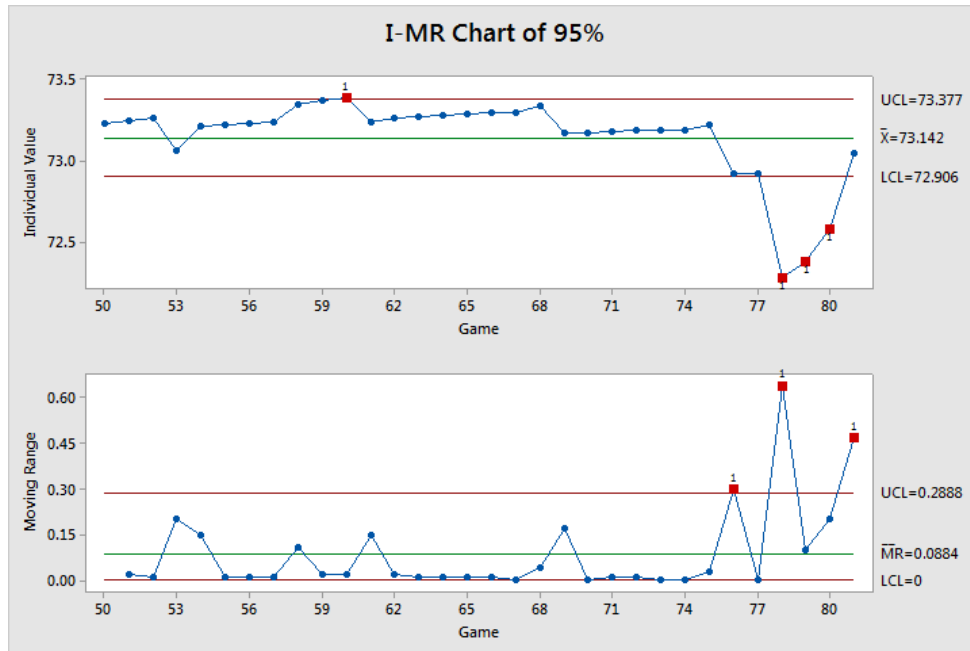
Figure 4. Control Chart of 73W Prediction

Though, at Game 61, Warriors lost to the worst West team Lakers. Even the prediction was still showing 95% confidence that Warriors could still win 73 games, how many decision makers would still stick with statistics and put the investment in at high risk/high return scenario? Based on the curve, games 65-68 might be good choices to purchase tickets instead. Though, ticket price had been gradually going up since Game 60. The ticker price had been increased from $231 to $275 at Game 68. Team decision might not be perfect at that time. Though, it was not a bad decision too. Team used money wisely. Team heard from one of classmates: their family purchased four lower seats at the expense of $8,000 on April 12. This learning had opened our eyes on how to integrate Predictive Statistics with Practical Risk Assessment. Mentors told us always: Statistics is all Risk Assessment and Money involved. Team started to understand what mentor was telling through the real example.

Control chart also indicated a very unstable period on Warriors' games on #78, #79, #80, and #81. Lost to Minnesota on Game #78 had almost killed Warriors' chance of winning 73 games. A BIG drop at Game #78 was shown in Figure 4. If looked at Warriors' games after #50, #78 game's loss to Minnesota had the biggest impact (probably the worst performance for the entire season). The coach and players might go back to understand more about this wild game in time, which might help players prevent this unacceptable failure in the future. Team won't ever know what Coach said to the players in their break room after Game #78 loss. Coach Kerr must do something very special in order to rebuild Warriors' team confidence to face the next three BIG games starting two days later. They would fly to Memphis the next morning. Team would bet, during the 2-3 hours flight, they must use every minute to discuss how to win the next three Big Games before going back to Bay Area to finish their 82nd game. As shown in Figure 4, there was an unbelievable coming back on Games 79-81 including playing the 2nd best team Spurs twice. Team was curious how coach and players could adjust their mindset and turn around quickly. This is a great team always do. Team was glad this was Warriors.

From the Moving Range Chart, team could observe this wild Up and Down pattern from Games 76-81. 3 out of control points were observed in 6 games. Team could imagine why coach Kerr mentioned that players were experiencing something they never been through. They were nervous and might over-do it. Fortunately, team had a great coach named Kerr, and a great two-times MVP leader named Curry. Kerr and Curry led a wonderful Warriors' team and made a 73-W history. NBA also announced Curry has won his 2nd MVP Award.

Even the 82nd game has been passed almost for a few months, team was so thankful that Warriors have made our life so beautiful in spring time. Since April 13, the IEOM team met each weekend to complete this STEM

Project. The predictive model may not be very accurate all the times, but the ending was surprisingly matching the prediction confidence made in early Feb. two months before the end result. The most important to team is that team believed in Statistics in daily life. Team started to believe more in Statistics and Probability, there is no absolute and perfect decision. There are always some degrees of risk associated with Statistics and Probability. Team should not give up the model or work just due to one bad game. Based on Statistical large sample principle, the model can be more accurate if team allowed enough time to monitor and modify the model continuously. The confidence on the risk management is how team can make a team decision together like purchasing tickets. Team has made two smart decisions based on predictive analytics: (1) team believed in Warriors, (2) team purchased tickets earlier. Thanks for IEOM for providing team this great opportunity to apply Statistics and Probability to support Warriors. STEM Project Contest opportunity certainly motivated team to explore the Statistics and Probability beyond AP Statistics Books. Team shared this story with friends and they were so curious about the Statistics. Team fully believes that team will continue learning Statistics and will come back Y2017. Warriors may break 73-W record next year, and team will definitely upgrade current model next year if Warriors can give us one more wonderful night in April, 2017. All of team members will remember this year's IEOM STEM Project Contest.

6 Future Work

Even team has completed this IEOM STEM Project and Warriors have set the 73-W record, team conducted another brainstorming session in early May on how to make model more accurately. Warriors, unfortunately lost to Lebron James' Cleveland Team in NBA Champion Games. In the off season, Warriors got Kevin Durant joining Warriors Team and which will make next 2016-2017 even more excited. Team has thought about the following opportunities:

- Did not consider the injury factor, especially on the key players like Curry. Team can certainly add this factor into model to more accurately reflect the real situations. Same for the Opponent Team.
- Did not consider the historical record between Warriors and the opponent team. They might have played one or two games in Warriors' first 50 games. This dependence information may be critical for certain match-up to upgrade the predictive model. Warriors may have a tough match-up with certain teams.
- Did not consider the trending or recent performance (last 10 games) into the modeling. Team may use the last 10 games of both teams to adjust the winning probability in real-time analysis.
- Did not consider the score difference of each game. The score difference of each game may be a more reliable source than just winning or losing a game. Score is a continuous variable and winning is a discrete or attribute variable. If team can convert the game outcome from Binomial distribution to Normal distribution, the predictive model may be more powerful and may require a smaller sample size, which means that team may be able to build a predictive model earlier.

7 Conclusions

Team has built a predictive model to simulate the probability whether 2015-2016 Warriors can break Chicago Bull's 1995-1996 72-Win record on April 13, 2016. Based on the first 50 games, the model can predict the remaining 32 games outcome reliably at more than 95% confidence. Based on the decision tree analysis, team was able to purchase the tickets of the final season game at 75% discount. From the Control Chart analysis, the unusual trending pattern can explain why Warriors have shown up and down performance under the pressure when approaching the season end to break the record. Warriors could overcome this challenging and made the wonderful history night on April 13, 2016. This project has utilized statistical data-driven methodology to uncover several interesting patterns.

Acknowledgements

would also like to thank my mom Chia Lin for her endless love to support my STEM Project. We had a good time together at the historical 82nd Game.

References

[1] 538 Prediction Site http://fivethirtyeight.com/features/warriors-record-73-wins/
[2] Peter Arcidiacono, Josh Kinsler, and Joseph Price. Productivity spillovers in team production: Evidence from professional basketball, 2014.
[3] Allan Maymin, Philip Maymin, and Eugene Shen. Nba chemistry: Positive and negative synergies in basketball. In 2012 MIT Sloan Sports Analytics Conference, 2012.
[4] Min-hwan Oh, Suraj Keshri, and Garud Iyengar. Graphical model for basketball match simulation. In 2015 MIT Sloan Sports Analytics Conference, 2015.
[5] Yahoo NBA Standing Record http://sports.yahoo.com/nba/standings/
[6] The Annals of Mathematical Statistics Vol. 16, No. 4 (Dec., 1945), pp. 319-329
[7] Handbook of Biological Statistics by John H. McDonald
[8] Introduction to Statistical Quality Control by Montgomery, Douglas (2005)

Biography

Mason Chen is a certified IASSC Black Belt. He has also certified IBM SPSS Statistics Certificate, IBM Modeler Data Analysis and Data Mining Certificates. Mason Chen is familiar with Lego Robotics/EV3, Six Sigma DMAIC, DMADOV, Lean Production, Minitab, SPSS Statistics, SPSS Modeler CRISP Data Mining, Applied Statistics, JAVA Programming, and ASQ Quality Engineering. He is the founder of Mason Chen Consulting which organization helps develop young kids on Big Data Analytics and STEM Projects.

Zonghuan (Jason) Li has certified SPSS statistics, SPSS Modeler data analyst and SPSS modeler data mining. Jason is familiar with Six Sigma DMAIC, SPSS Statistics, SPSS modeler CRISP Data Mining, and AP Statistics. Recently, he is learning "Computational Biology" which integrates Biology, Chemistry, Physics, Mathematics, Statistics, and Computer Science fields. At school, Jason is involved in Renaissance Leadership and many clubs. He is also on the JV basketball team.