Team Control Number

24463

For office use only	For office use only
T1 _____	F1 _____
T2 _____	F2 _____
T3 _____	F3 _____
T4 _____	F4 _____

Problem Chosen

B

2014 Mathematical Contest in Modeling (MCM) Summary Sheet

Summary:

We construct three models to study the definition of top 5 college coaches.

Model 1 is focus on the definition of a good coach in general. We argue that winning percentage and coaching career length are two so crucial factors that can measure one coach's success. We rank all the coaches by these two parameters separately. Then we compare them and get the intersection from which we pick out the top 38.2 percent (1-0.618) coaches as the "good coaches" we need.

Model 2 is focus on finding top 15 coaches of all time. We choose two extra factors, number of champions and performance in high-level games, for the next-step filtering. In order to make the source data internally consistent and each data type has the same kind of content and format, we choose one best standardization techniques to normalize the numbers. We calculate the four parameters of all coaches and realize that some of the parameters are more important than others somehow. So we introduce a function to calculate the weighted mean of these parameters via Analytical Hierarchy Process (AHP). Sorting the means, we get a new rank of "good coaches", from which we pick up the top 15 as the objects for further researching.

Model 3 intends to find the answer to "Who are top five greatest coaches of all-time" depended on more rigid parameters. We introduce the draft picks and wins to calculate each coach's input-output efficiency, which indicates the ability to realize more goals in less games, via Data Envelopment Analysis (DEA). Besides, we do researches about awards they have won to reflect their recognition by audience or experts. We sum up all these factors and give each one a score.

By ranking the score, we can see the final list of top 5 coaches of all time.

Considering the influence of time line horizon, we find the difficulty to win may change when time flows. We study the correlation coefficient between Strength of Schedule (SOS) and Winning Percentage so as to reflect the relationship between the difficulty of winning a game and the coach's ability to win a game. And we discuss each parameter to explain how we apply our model across both genders and all possible sports.

Who are Top 5 Coaches?

Abstract:

We construct three models to study the definition of top 5 college coaches.

Model 1 is focus on the definition of a good coach in general. We argue that winning percentage and coaching career length are two so crucial factors that can measure one coach's success. We rank all the coaches by these two parameters separately. Then we compare them and get the intersection from which we pick out the top 38.2 percent (1-0.618) coaches as the "good coaches" we need.

Model 2 is focus on finding top 15 coaches of all time. We choose two extra factors, number of champions and performance in high-level games, for the next-step filtering. In order to make the source data internally consistent and each data type has the same kind of content and format, we choose one best standardization techniques to normalize the numbers. We calculate the four parameters of all coaches and realize that some of the parameters are more important than others somehow. So we introduce a function to calculate the weighted mean of these parameters via Analytical Hierarchy Process (AHP). Sorting the means, we get a new rank of "good coaches", from which we pick up the top 15 as the objects for further researching.

Model 3 intends to find the answer to "Who are top five greatest coaches of all-time" depended on more rigid parameters. We introduce the draft picks and wins to calculate each coach's input-output efficiency, which indicates the ability to realize more goals in less games, via Data Envelopment Analysis (DEA). Besides, we do researches about awards they have won to reflect their recognition by audience or experts. We sum up all these factors and give each one a score.

By ranking the score, we can see the final list of top 5 coaches of all time.

Considering the influence of time line horizon, we find the difficulty to win may change when time flows. We study the correlation coefficient between Strength of Schedule (SOS) and Winning Percentage so as to reflect the relationship between the difficulty of winning a game and the coach's ability to win a game. And we discuss each parameter to explain how we apply our model across both genders and all possible sports.

Keywords:

Standardization, Analytical Hierarchy Process (AHP), Data envelopment analysis (DEA), Correlation Coefficient

Contents

Introduction

We focus on addressing modeling an optimal scheme so as to find the best college coach from all the time.

We can see that every year plenty of awards are given to reward the coaches performed excellent. For an instance, the **Adolph F. Rupp Cup**, an award given annually to the men's college basketball head coach in NCAA Division I competition, chooses someone *"who best exemplifies excellence in his dedication to the game of basketball and to his student athletes."* [2014]. However, when an award has been defined like this, the winner usually comes out as a result of voting, which means a high possibility of subjectivity. For another, the **Legends of Coaching Award**, when selecting the individual, considers a coach's character, success rate on the court, graduating rate of student athletes, his coaching philosophy and his identification with the goals of the John R. Wooden Award. [2014]. When it refers to success rate, which is a statistical index, we believe it can be one factor to define a great coach, while it cannot become the only factor to measure the greatness. Therefore, we seek a "fair" and "simple" model to find out who are the real great coaches of all time through necessary parameters.

We consider that several factors can be essential and decisive, at least most people believe that. The winning percentage is one undeniable index to judge one coach's ability. We usually believe that if one coach plays more games, he or she can be more experienced. The champions and awards one coach gets can measure his or her ability to win the games as well. Coaches are also responsible for the guidance of talents, so we also take drafts picks into consideration. Besides, we decide to judge one coach's performance in the high-level games, such as the winning percentage in the playoff games.

Conclude the factors above and we choose several some methods, such as Analytical Hierarchy Process (AHP) and Data Envelopment Analysis (DEA).

Assumptions

• We have accurate data about the coaches' win-loss-tie record, champions and other relevant factors

• We assume that the various awards mean the same in our research. We just measure one coach's honors through the number of awards.

• Our model is based on the data we can find, so we assume the other data not mentioned or some factors which cannot be quantified like temper or bad habits will not affect the results

Models

Model 1: Definition of Good Coaches

Definitions

Weighted Winning Percentage (WWP): We define a point system which is 2 for a win, 1 for a tie, and 0 for a loss. And the Weighted Winning Percentage can be calculated as follows:

$$\text{WWP}=\frac{(1\times\text{Ties})+(2\times\text{Wins})}{2\times\text{Games Played}}$$

Coaching Career Length (CCL): We describe a coach's total games as his or her coaching career length.

Picking out Good Coaches

According to the statistics we have, no matter which kind of sports, it has a large quantity of data which bring trouble in finding the greatest coaches directly. In that way, we need select "Good Coaches" through two simple but undeniable parameters, winning percentage and coaching career length. Therefore, we make sure our "good coaches" list can be efficient when we find good coaches who can win games as many as possible and exclude the high winning percentage caused by a few games played in the meantime. Then a "good coach" is a good winner as well as a well-experienced teacher.

By sorting these two parameters in descending order separately, we get the rankings of all coaches' winning percentage and coaching length. We choose the top 38.2 percent (1-0.618) as two sets named "good winning percentage" and "good coaching career length". Compare these two sets and get the intersection which we call it "Good Coach".

Model 2: Top 15 Coaches

Definitions

CCL': A standardized index of *CCL*.
Champions Experience (CE): We define the number of champions as one coach's Champions Experience. For different sports, we calculate the different levels of champions and give them weighted ratios.
CE': A standardized index of *CE*.
High-level Games Performance (HGP): We select the high-level games of each sport, such as Bowl Games of NCAA Football, to judge the coach's performance when he or she met high-level adversaries in high-intensity games
HGP': A standardized index of *HGP*.
Greatness (G): We define the weighted mean of *WWP, CCL', CE'* and *HGP'* as one coach's greatness.

Best Data Standardization Techniques

In order to make the source data internally consistent and each data type has the same kind of content and format, here we prepare three standardization techniques:

• Each variable in the data set is recalculated as $\frac{V - V_{\min}}{V_{\max} - V_{\min}}$, where V is the value of the variable in the original data set.

• Each variable is recalculated as $\frac{\log_{10} V}{\log_{10} V_{\max}}$

• Each variable is recalculated as $\frac{\arctan(V) \times 2}{\pi}$

We analyze the data after standardization and figure out that there are some flaws in the No.2 and No.3 standardization plan.

When we recalculate each variable as $\frac{\log_{10} V}{\log_{10} V_{\max}}$, we can see that when V is between 0 and 1 the result will be a negative number which does not meet the requirement. Here is the graph for $\log_{10} V$ (**Figure 1**)

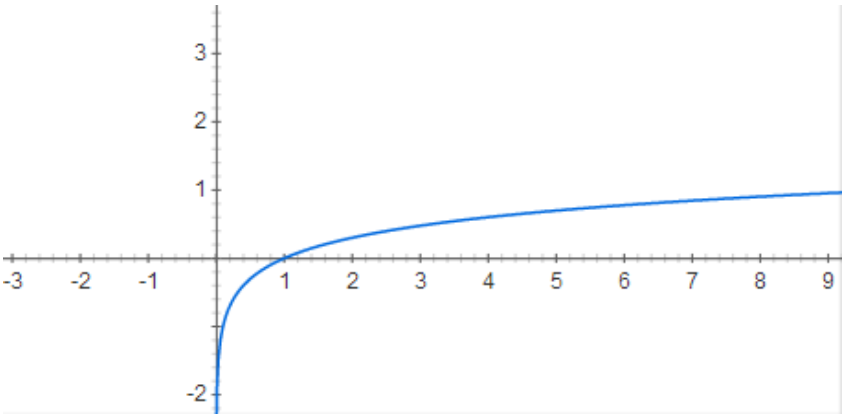


Figure 1

When we recalculate each variable as $\frac{\arctan(V) \times 2}{\pi}$, we can see that when V is a large number, such as the number of one coach's games played which normally is bigger than 100, there are tiny differences between the results. Here is the graph for $\frac{\arctan(V) \times 2}{\pi}$ (Figure 2)

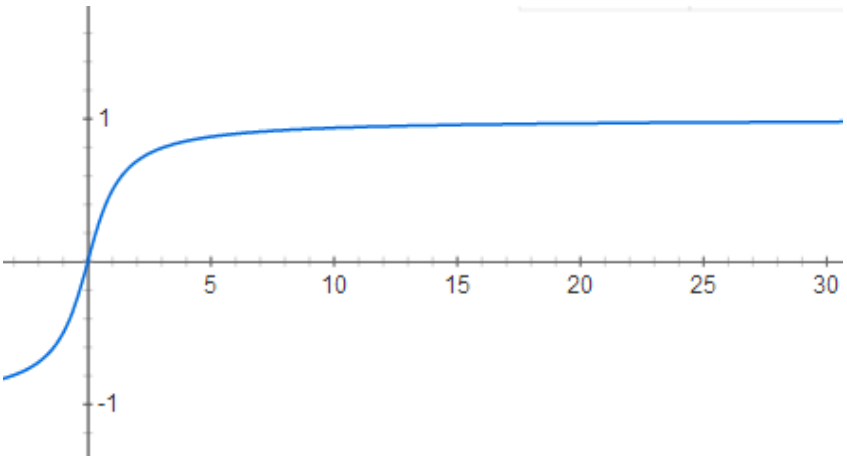


Figure 2

Therefore, we choose the first standardization plan that each variable set is recalculated as $\frac{V-V_{\min}}{V_{\max}-V_{\min}}$. In this way we have four parameters, *WWP*, *CCL*, *CE* and *HGP* for the further ranking.

Construction

Given a specific coach, we calculate the four parameters of him or her and compare them with all other coaches' parameters. We realize that some of the parameters are more important than others somehow. So in order to make our model become more accurate and convincible, we introduce a weighted mean of these parameters X, with

$$\overline{X} = \sum_{i=1}^{4} w_i . X_i$$

We determine the weights via the Analytical Hierarchy Process (AHP) [Saaty 1982]. We build a 4×4 matrix reciprocal matrix by pair comparison:

	WWP	CE	CCL	HGP
WWP	1	2	4	5
CE	$\frac{1}{2}$	1	2	3
CCL	$\frac{1}{4}$	$\frac{1}{2}$	1	2
HGP	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{1}{2}$	1

The meaning of the number in each cell is explained in **Table 2**. The numbers themselves are based on our own subjective decisions.

Table 2.
The multiplication table of D_{10}.

Intensity of Value	Interpretation
1	Requirements *i* and *j* have equal value.
3	Requirement *i* has a slightly higher value than *j*.
5	Requirement *i* has a strongly higher value than *j*.
7	Requirement *i* has a very strongly higher value than *j*.
9	Requirement *i* has an absolutely higher value than *j*.
2, 4, 6, 8	Intermediate scales between two adjacent judgments.
Reciprocals	Requirement *i* has a *lower* value than *j*.

Then we input the matrix into a Matlab program that calculates the weight w_i of each factor, as given in **Table 3**.

Table 3.
AHP-derived weights.

Factor	WWP	CE’	CCL’	HGP’
Weight	0.4611	0.2653	0.1424	0.1312

We test the consistency of the preferences for this instance of the AHP. For good consistency [Alonso and Lamata 2006, 446–447]:
• The principal eigenvalue λ_{max} of the matrix should be close to the number n of alternatives, here 4; we get $\lambda_{max} = 3.9314$.
• The consistency index CI = $(\lambda_{max} - n)/(n - 1)$ should be close to 0; we get CI = -0.0229.
• The consistency ratio CR = CI/RI (where RI is the average value of CI for random matrices) should be less than 0.01; we get CR = -0.0257.

Therefore, our decision method displays perfectly acceptable consistency and the weights are reasonable.

Results

Our parameter used to pick out top 15 coaches is calculated as follows:

$$\overline{X} = \sum_{i=1}^{4} w_i \cdot X_i = 0.4611 \times \mathrm{WWP} + 0.2653 \times \mathrm{CE} + 0.1424 \times \mathrm{CCL} + 0.1312 \times \mathrm{HGP}$$

Model 3: Top 5 Greatest Coaches

Definitions

Ability to Win (AW): We define the number of one coach’s the winning games as his or her ability to win.

Draft Picks (DP): We count the draft picks of one coach so as to reflect his or her ability to guide talents being a pro.
Efficiency (E): We calculate the efficiency of one coach by Data Envelopment Analysis. [1978]
Honors (H): We standardize the number of awards one coach have got his or her whole life and call it "Honors".
Score (S): We give everyone a score rely on one's Efficiency and Greatness.

Introduction

Now we have the top 15 coaches who have been well-known and recognized as "great coaches" or "legends". However, we aim at the top 5 greatest coaches so we have to construct a more rigid model with some other data. We choose CCL, AW, CE, DP and HGP to do the Data Envelopment Analysis.

Data Envelopment Analysis (DEA)

Data Envelopment Analysis (DEA) is a tool for measuring and comparing the relative efficiency of decision making units (DMU). The DMUs are usually described by several inputs spent to produce several outputs. Let us consider the set E of n decision making units E= {DMU_1, DMU_2, … , DMU_n }. Each of units produce r outputs, and for their production m inputs are spent. [2007] Here we define CCL as the only input and AW, CE, DP and HGP as outputs. And we get:

		DMU_1	…	DMU_i	…	DMU_n
Inputs	CCL	x_1	…	x_i	…	x_n
Outputs	AW	y_{11}	…	y_{1i}	…	y_{1n}
	CE	y_{21}	…	y_{2i}	…	y_{2n}
	DP	y_{31}	…	y_{3i}	…	y_{3n}
	HGP	y_{41}	…	y_{4i}	…	y_{4n}

We can write the outputs of DMU_i as:

$$y_i = (y_{1i}, y_{2i}, y_{3i}, y_{4i})^T$$

And we get:

	DMU_1	…	DMU_i	…	DMU_n
Inputs	x_1	…	x_i	…	x_n
Outputs	y_1	…	y_i	…	y_n

Then we write the Inputs and Outputs as:

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}$$

And we define the weights of Outputs as:

$$u = (u_1, u_2, u_3, u_4)^T$$

In this way, we can use the ratio of Outputs and Inputs to reflect the efficiency of DMU:

$$E_{ii} = \frac{y_i^T u}{x_i^T v}$$

Then we use a Matlab program to calculate the best weights of Output and E_{ii}, which can transfer to the index "*E*". [2002]

Honors

We think that the honors a coach has got indicate his or her recognition among people. In spite of the variation of awards, we still count the total number of his or her prizes. After we get the numbers, it is necessary to standardize them for the calculation of *S*. Here we take the "z-scores" transformation for standardization [2014]:

$$H = \frac{N - \overline{N}}{\sigma}$$

(N is the number of awards; $\overline{N}$ is the mean; σ is the standard deviation.)

Efficiency, Greatness, and Honors

Now we have *E, G* and *H*, we use these three types of data to give top 15 coaches a score which and drastically judge if a coach is the greatest one. The score is calculated as:

$$S = 20 \times E + 80 \times G + H$$

We rank the scores and get the top 5 greatest coaches.

Analysis

Time Line Horizon and Greatest Coaches

We consider that most of the parameters are not affected by the time line horizon except for the difficulty to win a game. The difficulty to win a game is changing by time and the changes of opponents. So we introduce a parameter named **strength of schedule** (**SOS**).

In sports, **strength of schedule** (**SOS**) refers to the difficulty or ease of a team's/person's opponent as compared to other teams/persons. This is especially important if teams in a league do not play each other the same number of times. [2014] The SOS was calculated as follows:

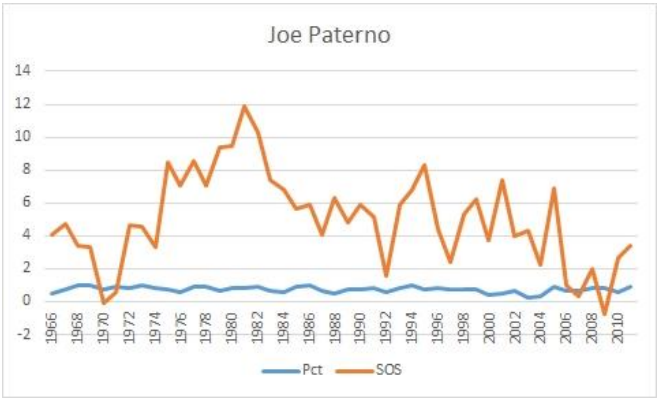
$$SOS = \frac{2(OR)+(OOR)}{3}$$

OR is the opponents' record, while OOR is the opponents' opponents record [2014]

We calculate the correlation coefficient between SOS and Winning Percentage so as to reflect the relationship between the difficulty of winning a game and the coach's ability to win a game. Take football as an example first and the result is as follows:

Joe Paterno:

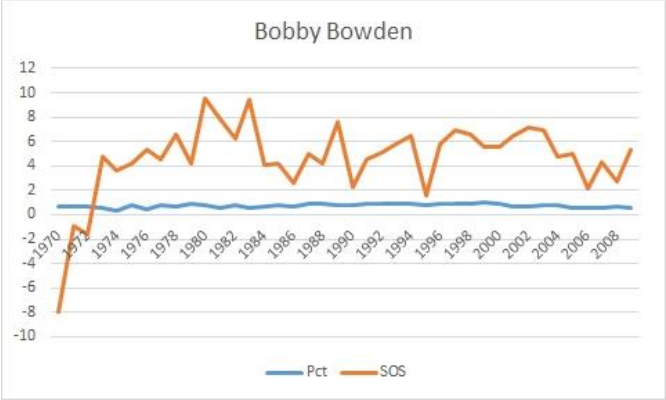
Year	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976
Pct	0.5	0.773	1	1	0.7	0.917	0.833	1	0.833	0.75	0.583
SOS	4.06	4.71	3.37	3.29	-0.06	0.59	4.64	4.53	3.3	8.47	7.08
Year	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
Pct	0.917	0.917	0.667	0.833	0.833	0.917	0.654	0.545	0.917	1	0.667
SOS	8.52	7.08	9.41	9.44	11.86	10.27	7.35	6.78	5.62	5.91	4.09
Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Pct	0.708	0.75	0.846	0.583	0.833	1	0.75	0.846	0.75	0.75	0.769
SOS	4.83	5.87	5.13	1.58	5.93	6.77	8.28	4.43	2.4	5.29	6.25
Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	
Pct	0.417	0.455	0.692	0.25	0.364	0.917	0.692	0.5435	0.846	0.846	
SOS	3.71	7.39	3.98	4.35	2.24	6.91	0.97	3.9633	1.97	-0.75	



R=0.1134

Bobby Bowden:

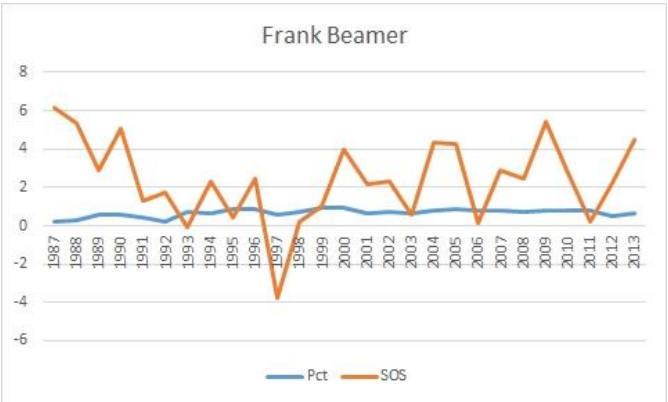
Year	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
Pct	0.727	0.636	0.667	0.545	0.364	0.75	0.455	0.833	0.727	0.917
SOS	-7.96	-0.88	-1.62	4.81	3.65	4.23	5.35	4.56	6.59	4.24
Year	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Pct	0.833	0.545	0.75	0.583	0.667	0.75	0.625	0.917	0.917	0.833
SOS	9.53	7.8	6.23	9.38	4.1	4.15	2.6	5	4.24	7.64
Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Pct	0.846	0.917	0.923	0.875	0.833	0.917	0.917	0.846	1	0.846
SOS	4.59	5.08	5.76	6.48	1.61	5.79	6.88	6.57	5.58	5.59
Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	
Pct	0.667	0.643	0.769	0.75	0.615	0.538	0.538	0.692	0.538	
SOS	6.45	7.15	6.9	4.75	5.05	2.19	4.33	2.67	5.38	



R=0.1211

Frank Beamer:

Year	1987	1988	1989	1990	1991	1992	1993	1994	1995
Pct	0.182	0.273	0.591	0.545	0.455	0.227	0.75	0.667	0.833
SOS	6.16	5.39	2.91	5.1	1.3	1.7	-0.04	2.33	0.46
Year	1996	1997	1998	1999	2000	2001	2002	2003	2004
Pct	0.833	0.583	0.75	0.917	0.917	0.667	0.714	0.615	0.769
SOS	2.43	-3.79	0.19	0.99	4	2.16	2.31	0.6	4.33
Year	2005	2006	2007	2008	2009	2010	2011	2012	2013
Pct	0.846	0.769	0.786	0.714	0.769	0.786	0.786	0.538	0.615
SOS	4.27	0.16	2.86	2.43	5.42	2.74	0.21	2.21	4.49

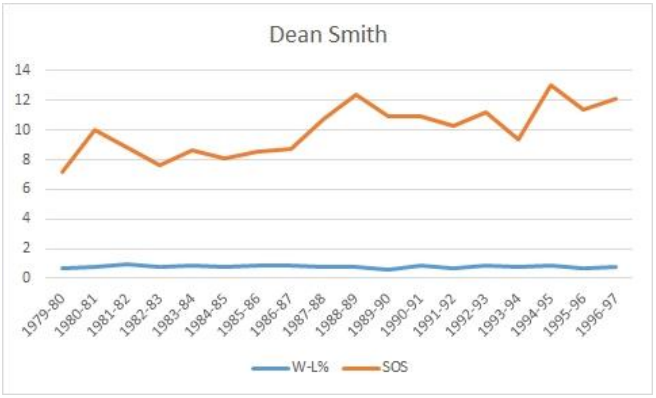


R=0.2282

Then we take basketball as an example and the result is as follows:

Dean Smith:

Season	1979-80	1980-81	1981-82	1982-83	1983-84	1984-85	1985-86	1986-87	1987-88
W-L%	0.724	0.784	0.941	0.778	0.903	0.75	0.824	0.889	0.794
SOS	7.18	10.05	8.87	7.68	8.6	8.12	8.51	8.71	10.72
Season	1988-89	1989-90	1990-91	1991-92	1992-93	1993-94	1994-95	1995-96	1996-97
W-L%	0.784	0.618	0.829	0.697	0.895	0.8	0.824	0.656	0.8
SOS	12.41	10.97	10.91	10.29	11.25	9.36	13.01	11.43	12.09



R=0.1306

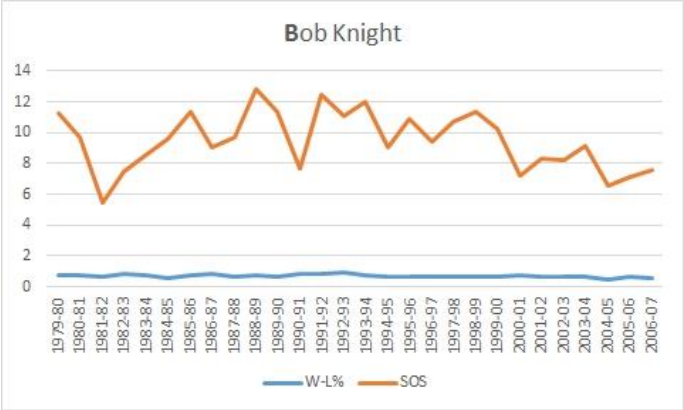
Mike Krzyzewski:

Season	1983-84	1984-85	1985-86	1986-87	1987-88	1988-89	1989-90	1990-91	1991-92	1992-93	1993-94
W-L%	0.706	0.742	0.925	0.727	0.8	0.778	0.763	0.821	0.944	0.75	0.824
SOS	6.92	7.38	9.5	7.16	8.38	8.19	8.81	10.62	10.3	10.76	8.92
Season	1994-95	1995-96	1996-97	1997-98	1998-99	1999-00	2000-01	2001-02	2002-03	2003-04	2004-05
W-L%	0.75	0.581	0.727	0.889	0.949	0.853	0.897	0.886	0.788	0.838	0.818
SOS	11.89	10.48	9.27	9.33	10.14	8.88	11.98	9.16	8.38	10.86	10.04
Season	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14		
W-L%	0.889	0.667	0.824	0.811	0.875	0.865	0.794	0.833	0.792		
SOS	8.77	10.49	9.28	9.07	9.21	8.32	8.4	9.54	7.94		

R=0.145

Bob Knight:

Season	1979-80	1980-81	1981-82	1982-83	1983-84	1984-85	1985-86	1986-87	1987-88	1988-89
W-L%	0.724	0.743	0.655	0.8	0.71	0.576	0.724	0.882	0.655	0.771
SOS	11.25	9.66	5.42	7.45	8.63	9.6	11.41	9.1	9.66	12.84
Season	1989-90	1990-91	1991-92	1992-93	1993-94	1994-95	1995-96	1996-97	1997-98	1998-99
W-L%	0.621	0.853	0.794	0.886	0.7	0.613	0.613	0.667	0.625	0.676
SOS	11.37	7.68	12.44	11.09	12.05	9.05	10.87	9.4	10.71	11.36
Season	1999-00	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07		
W-L%	0.69	0.719	0.629	0.676	0.667	0.469	0.618	0.6		
SOS	10.22	7.22	8.29	8.21	9.11	6.56	7.16	7.62		



R=0.2627

Several authors have offered guidelines for the interpretation of a correlation coefficient. [1988] When r is between 0.1 and 0.3, we believe that two factors have a small relation. In this way, we know that the coach's ability to win has tiny relations with the difficulty to win.

This result implies that a great coach will not be influenced by the changes of opponents. In other words, it makes no difference which timeline horizon line we use in our analysis.

The Applicability of the Model across Both Genders and All Possible Sports

Greatness and Genders:

Since a coach's task is to guide and train athletes, we assume that this job which cares about brain and method has nothing to do with genders of the coaches.

No matter the winning percentage, champions, or career length makes no differences between male and female coaches. The only factor that may cause errors is the awards. Due to the fact that some awards are given separately to male and female, there are distinctions led by the competitive pressure. However, a few awards are set only for the male which seems unfair to the female coaches. We balance these two situations and conclude that genders cause tiny differences which can be hardly taken into account.

Greatness and Sports:

We realize different sports may cause great differences, so it is necessary to discuss every factor individually as follows:

• *WWP*: Some sports like basketball do not have a tie game, while other sports like soccer always end up with a tie. We define the point system which is 2 for a win, 1 for a tie, and 0 for a loss. And calculating winning percentage as $\text{WPP}=\frac{(1\times\text{Ties})+(2\times\text{Wins})}{2\times\text{Games Played}}$ makes sure there are no differences no matter whether there is a tie or not.

• *CE*: Different sports have different champions. Almost every sport has the champion of the league annually. Besides, some sports like college basketball which has Conference Champion, NCAA Tournament Champion, and NCAA (national) Champion need to be considered respectively. We determine the weights of the different levels of champions via the Analytical Hierarchy Process. Then we calculate and get the result *CE*.

• *CCL*: We consider that the kind of sports has no effect on coaching career length.

• *HGP*: We consider that the kind of sports makes no differences when calculating high-level game performance. However, we need to define high-level games of various sports. For instance, we select the Bowl Games as the high-level games when we choose the greatest football college coaches.

• *AW*: We consider that the kind of sports has no effect on one coach's ability to win.

• *DP*: Most college sports have drafts to send their students to the professional league. However, some specific sports do not have drafts. Then we ignore Model 2 and get the *S* of the coach directly.

• *H*: Each sport has its own awards to reward the coaches, so the kind of sports has no effect on one coach's honors.

Results:

Football:

• Model 1:

After comparing these two sets and get the intersection. We find that there are 358

NAME	START YEAR	END YEAR	SCHOOL	CCL	CCL'	WIN(AW)	LOSE	TIE	WWP	CE	CE'	HGP'	G
Joe Paterno	1966	2011	Penn State	548	1.00000	409	136	3	0.749	37	1.00000	0.662	0.83992
Bobby Bowden	1970	2009	Florida State	485	0.87221	357	124	4	0.74	33	0.89189	0.682	0.79151
Bear Bryant	1945	1982	Alabama	425	0.75051	323	85	17	0.78	29	0.78378	0.552	0.74689
Tom Osborne	1973	1997	Nebraska	307	0.51116	255	49	3	0.836	25	0.67568	0.48	0.70050
Lou Holtz	1969	2004	South Carolina	388	0.67546	249	132	7	0.651	22	0.59459	0.591	0.63165
Mack Brown	1985	2013	Texas	356	0.61055	238	117	1	0.67	21	0.56757	0.619	0.62767
LaVell Edwards	1972	2000	Brigham Young	361	0.62069	257	101	3	0.716	22	0.59459	0.341	0.62102
Steve Spurrier	1987	2013	South Carolina	300	0.49696	219	79	2	0.733	19	0.51351	0.474	0.60718
John Vaught	1947	1973	Mississippi	263	0.42191	190	61	12	0.745	18	0.48649	0.556	0.60561
Vince Dooley	1964	1988	Georgia	288	0.47262	201	77	10	0.715	20	0.54054	0.45	0.59943
Barry Switzer	1973	1988	Oklahoma	190	0.27383	157	29	4	0.837	13	0.35135	0.615	0.59884
Frank Beamer	1987	2013	Virginia Tech	335	0.56795	224	109	2	0.672	21	0.56757	0.429	0.59760
Bo Schembechler	1963	1989	Michigan	307	0.51116	234	65	8	0.775	17	0.45946	0.294	0.59061
Urban Meyer	2001	2013	Ohio State	153	0.19878	128	25	0	0.837	8	0.21622	0.875	0.58641
Darrell Royal	1954	1976	Texas	249	0.39351	184	60	5	0.749	16	0.43243	0.531	0.58579

good coaches who meet our requirement.

• Model 2:

We calculate the four parameters, *CCL', CE', WWP'* and *HGP'*, and the result, *G*. We pick up the top 15 and list them as follows:

• Model 3:

We calculate the parameter *E* and *H*, then we get the score *S*. The results is listed and showed as follows:

NAME	G	DP	E	G*0.8+E*0.2	AWARDS	H	S
Joe Paterno	0.83992	30	0.98570	0.86907	23	2.20	89.11
Bobby Bowden	0.79151	24	1.00000	0.83321	3	-1.26	82.06
Bear Bryant	0.74689	21	0.91910	0.78133	15	0.82	78.95
Tom Osborne	0.70050	25	1.00000	0.76040	15	0.82	76.86
Lou Holtz	0.63165	7	1.00000	0.70532	13	0.47	71.00
Mack Brown	0.62767	22	1.00000	0.70213	2	-1.43	68.78
LaVell Edwards	0.62102	12	0.85280	0.66738	17	1.16	67.90
John Vaught	0.60561	22	0.91630	0.66775	13	0.47	67.25
Barry Switzer	0.59884	16	0.87690	0.65445	9	-0.22	65.23
Frank Beamer	0.59760	5	0.90630	0.65934	6	-0.74	65.19
Darrell Royal	0.58579	13	0.88770	0.64617	8	-0.39	64.22
Steve Spurrier	0.60718	16	0.79190	0.64412	9	-0.22	64.19
Vince Dooley	0.59943	20	0.84210	0.64797	4	-1.08	63.71
Urban Meyer	0.58641	6	0.80430	0.62999	13	0.47	63.47
Bo Schembechle	0.59061	9	0.85490	0.64347	4	-1.08	63.26

• Top 5 Football College Coaches:

Joe Paterno, Bobby Bowden, Bear Bryant, Tom Osborne, Urban Meyer.

Basketball:

• Model 1:

After comparing these two sets and get the intersection. We find that there are 579 good coaches who meet our requirement.

• Model 2:

We calculate the four parameters, *CCL', CE', WWP'* and *HGP'*, and the result, *G*. We pick up the top 15 and list them as follows:

NAME	START YEAR	END YEAR	SCHOOL	CCL	CCL'	(WIN)AW	LOSE	WWP	HGP'	CE'	G
John Wooden	1947	1975	UCLA	826	0.60264	664	162	0.804	0.88333	0.83403	0.78427
Adolph Rupp	1931	1972	Kentucky	1066	0.81410	876	190	0.822	0.54167	0.58074	0.74005
Mike Krzyzewski	1976	2014	Duke	1277	1.00000	975	302	0.764	0.92917	0.44531	0.73099
Dean Smith	1962	1997	North Carolina	1133	0.87313	879	254	0.776	0.91250	0.34791	0.69362
Jim Calhoun	1973	2012	Connecticut	1259	0.98414	877	382	0.697	0.62917	0.29636	0.63997
Roy Williams	1989	2014	North Carolina	902	0.66960	715	187	0.793	0.50000	0.18770	0.62815
Bob Knight	1966	2008	Texas Tech	1273	0.99648	899	374	0.706	0.54583	0.30269	0.63801
Jim Boeheim	1977	2014	Syracuse	1256	0.98150	942	314	0.75	0.40000	0.26179	0.62198
Jerry Tarkanian	1970	2002	Fresno State	963	0.72335	761	202	0.79	0.44167	0.40436	0.64878
Rick Pitino	1979	2014	Louisville	920	0.68546	681	239	0.74	0.54583	0.19968	0.59844
Denny Crum	1972	2001	Louisville	970	0.72952	675	295	0.696	0.58750	0.27031	0.60476
Lute Olson	1974	2007	Arizona	1061	0.80969	776	285	0.731	0.36667	0.24277	0.57922
Phog Allen	1906	1956	entral Missour	978	0.73656	719	259	0.735	0.56667	0.32109	0.59871
Hank Iba	1930	1970	Oklahoma State	1085	0.83084	752	333	0.693	0.22083	0.27300	0.58475
John Calipari	1989	2014	Kentucky	756	0.54097	585	171	0.774	0.31667	0.26668	0.58154

• Model 3:

We calculate the parameter *E* and *H*, then we get the score *S*. The results is listed and showed as follows:

NAME	CCL	CCL'	(WIN)AW	LOSE	WWP	HGP'	CE'	G	DP	E	Awards	S
John Wooden	826	0.60264	664	162	0.804	0.88333	0.83403	0.78427	36	1.0000	17	83.89054
Adolph Rupp	1066	0.81410	876	190	0.822	0.54167	0.58074	0.74005	27	1.0000	14	79.83627
Mike Krzyzewsk	1277	1.00000	975	302	0.764	0.92917	0.44531	0.73099	44	0.9479	12	77.72523
Dean Smith	1133	0.87313	879	254	0.776	0.91250	0.34791	0.69362	71	1.0000	18	76.81117
Roy Williams	902	0.66960	715	187	0.793	0.62917	0.29636	0.63997	27	0.9857	17	72.06085
Jim Boeheim	1256	0.98150	942	314	0.75	0.50000	0.18770	0.62815	34	0.9231	13	69.17372
Bob Knight	1273	0.99648	899	374	0.706	0.54583	0.30269	0.63801	40	0.8705	14	69.08292
Jerry Tarkania	963	0.72335	761	202	0.79	0.40000	0.26179	0.62198	56	1.0000	2	68.32294
Jim Calhoun	1259	0.98414	877	382	0.697	0.44167	0.40436	0.64878	32	0.8522	3	67.68348
Lute Olson	1061	0.80969	776	285	0.731	0.54583	0.19968	0.59844	47	0.9199	13	66.73294
Rick Pitino	920	0.68546	681	239	0.74	0.58750	0.27031	0.60476	18	0.9121	7	66.04930
John Calipari	756	0.54097	585	171	0.774	0.36667	0.24277	0.57922	20	0.9446	15	66.03449
Denny Crum	970	0.72952	675	295	0.696	0.56667	0.32109	0.59871	35	0.8978	6	65.10651
Phog Allen	978	0.73656	719	259	0.735	0.22083	0.27300	0.58475	19	0.8946	2	63.23680
Hank Iba	1085	0.83084	752	333	0.693	0.31667	0.26668	0.58154	14	0.8434	2	61.95637

• Top 5 Basketball College Coaches:

John Wooden, Adolph Rupp, Mike Krzyzewsk, Dean Smith, Roy Williams.

Softball:

• Model 1:

After comparing these two sets and get the intersection. We find that there are 52 good coaches who meet our requirement.

• Model 2&3:

We calculate the four parameters, *CCL', CE'* and *WWP',* and the result, *S.* We pick up the top 15 and list them as follows:

Name	Years	CCL	CCL'	Wins(AW)	Losses	Ties	WWP	CE	CE'	S
Mike Candrea	26	1667	0.65535	1343	322	2	0.806	8	0.57143	74.253
Margie Wright	33	2002	1.00000	1457	542	3	0.729	0	0.00000	72.267
Yvette Girouard	31	1707	0.69650	1285	421	1	0.753	2	0.14286	67.452
Carol Hutchins	30	1777	0.76852	1325	448	4	0.747	0	0.00000	67.435
Gayle Blevins	31	1838	0.83128	1245	588	5	0.679	0	0.00000	64.724
JoAnne Graf	25	1617	0.60391	1186	425	6	0.735	0	0.00000	62.418
Elaine Sortino	34	1699	0.68827	1185	508	6	0.699	0	0.00000	62.304
Sue Enquist	19	1063	0.03395	887	175	1	0.835	11	0.78571	62.293
Eugene Lenti	32	1746	0.73663	1170	570	6	0.672	0	0.00000	61.833
Diane Ninemire	26	1698	0.68724	1175	523	0	0.692	0	0.00000	61.832
Ralph Weekly	26	1490	0.47325	1119	369	2	0.752	2	0.14286	61.622
Frank Cheek	25	1512	0.49588	1147	363	2	0.759	0	0.00000	61.157
Les Stuedeman	17	1113	0.08539	847	265	1	0.761	14	1.00000	61.151
Kathy Welter	31	1709	0.69856	1129	573	7	0.663	0	0.00000	60.277
Chris Bellotto	29	1517	0.50103	1123	394	0	0.74	0	0.00000	60.080

• Top 5 Softball College Coaches:

Mike Candrea, Margie Wright, Yvette Girouard, Carol Hutchins, Gayle Blevins

Strengths and Weaknesses

Strengths:

• We consider all possible essential factors like winning percentage which affect the great-ness of a coach. In this way this model can be more acceptable and fit most people's views.
• Our model considers many parameters and exact numbers, which means the judgment is objective.
• Our model is constructed step by step, which simplifies the process of selecting the greatest coaches
• Analytical Hierarchy Process (AHP) and Data Envelopment Analysis (DEA) have been perfectly used in our models, and the simulation results are consistent with the reality.

Weaknesses:

• Some special data cannot be found, and it makes that we have to do some proper assumption before the solution of our models. A more abundant data resource can guarantee a better result in our models.
• Our model is based on large quantities of data. For this reason, the model is more convincible when the data bases need to be relatively large.

Future Work

Keyword Frequency

After we rank the scores, we come up with an idea that keyword frequency can be another parameter reflecting how popular on coach is. We use "www.google.com" as our search engine to figure out the keyword frequency. We input the name of a coach and record the number of results related to it. Since the Internet starts from 1969, it is highly possible to be not accurate enough. So we just count it and provide it as a reference resources.

References

[1] Alonso, Jos´e Antonio, and Ma [Mar´ıa] Teresa Lamata. 2006. Consistency in the analytic hierarchy process: A new approach. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 14 (4): 445–459. http://hera.ugr.es/doi/16515833.pdf .
[2] Peng Yuwei, Xu Xiaozhan , and Wu Shouxian. 2002. *MATLAB Programs for DEA.*
[3] Saaty, Thomas L. 1982. *Strategy and Organization, The Analytical Hierarchy Process for Decisions in a Complex World*. Belmont, CA: Lifetime Learning Pub.

[4] Wikipedia. 2014. Adolph Rupp Cup. http://en.wikipedia.org/wiki/ Adolph_Rupp_Cup
[5] Wikipedia. 2014. Standard Score. http://en.wikipedia.org/wiki/ Standard_score
[6] Wikipedia. 2014. Strength of schedule. http://en.wikipedia.org/wiki/ Strength_of_schedule
[7] John R. Wooden Award. 2014. *Legends of Coaching*. http://www.woodenaward.com
[8] Kreyszig, E (fourth edition 1979). *Applied Mathematics*. Wiley Press.
[9] Josef Jablonsky. 2007. *Measuring the efficiency of production units by AHP models*.
[10] 2014, http://www.sports-reference.com/cbb/; 2014, http://www.sports-reference.com/cfb/.
[11] Cohen. J, 1988, *Statistical power analysis for the behavioral sciences*.

Appendix

Article for *Sports Illustrated*

Who can be Top 5 College Coaches

As Anthony Davis was taken with the top overall pick in 2012 NBA draft, as Louisville won the 75th edition of the NCAA men's basketball championship, perhaps lots of people did not notice the importance of their head coaches, or someone just simply considered their coaches as greatest college coaches of all time. Some people may ask, “What factors can we use to evaluate if a college coach deserve the title of greatness?” or “Who is the greatest coach in the history exactly?” Today we will discuss it with a relatively scientific method.

One friend told me that Pete Carroll, who these days was the head coach of Seattle Seahawks, was likely to be one of the greatest college coaches for the reason that his fantastic NCAA Football career record brought him a considerable contract of NFL, let alone he just led his team to win the 2014 Super Bowl Champion. All these fames seem to be enough to justify his greatness. However, I have to see it as subjective judgment before any statistical result comes out.

Recently, I, with my two partners, designed a method to solve this problem. It is a simple model but valid enough because it is based on a huge number of objective statistics. What we did was as follows:

Step1. Thanks to the great quantity of coaches, we decided to define what a good coach is. We argued that winning percentage and coaching career length are two so important factors that can measure one coach’s success. So we ranked all the coaches by these two parameters separately. Then we compared them and made the top 38.2 percent (1-0.618) coaches become the “good coaches” we needed.

Step 2. After we got our “good coaches”, we decided to use some greater things to find top 15 coaches of all time. We chose two extra factors, number of champions and performance in high-level games, for the next-step filtering. We calculated the four parameters of all coaches and realized that some of the parameters are more important

than others somehow. So we introduced a function to calculate the weighted mean of these parameters. Sorting the means, we got a new rank of "good coaches", from which we picked up the top 15 as the objects for further researching.

Step 3. When we got the list of these 15 coaches, we realized that they are all the coaches that people usually called as "great coaches" or "legends". Hence we have to depend on more rigid parameters to judge them finding the answer to "Who are top five greatest coaches of all-time". We introduced the draft picks and wins to calculate each coach's input-output efficiency which indicates the ability to realize more goals in less games. Besides, we did researches about awards they had won to reflect their recognition by audience or experts.

Step 4. Finally, we summed up all these factors and gave every coach a score. By ranking the score, we could see the final list of the greatest coaches of all time.

We tested our model with three sports: football, basketball and softball. And we listed the top five college coaches as follows:

	Football
1	Joe Paterno
2	Bobby Bowden
3	Bear Bryant
4	Tom Osborne
5	Urban Meyer

	Basketball
1	John Wooden
2	Adolph Rupp
3	Mike Krzyzewsk
4	Dean Smith
5	Roy Williams

	Softball
1	Mike Candrea
2	Margie Wright
3	Yvette Girouard
4	Carol Hutchins
5	Gayle Blevins

Do you satisfy with this list? Is there someone you wish to see on the list? If you have enough reasons to convince us that someone should be on the list but not, you can contact us and share your opinions with us. We are looking forward to further researching and optimizing the model to the largest extent.