May the best (statistically chosen) team win!

Danielle Pope

The Burning Question:

What does the Pythagorean Expectation tell us, and how can the Pythagorean Expectation be improved?

Pythagorean Expectation of Winning Percentage

- Formula invented by Bill James to estimate how many games a baseball team "should" have won based on the number of runs they scored and allowed.
- x^2/(x^2+y^2) where x=Runs Scored, y=Runs Allowed
- Name derived from the formula's resemblance to Pythagoras' formula to compute the length of the hypotenuse of a triangle from the lengths of its other two sides.

Example of Pythagorean Expectation Formula

2001 Arizona Diamondbacks

X=Runs Scored= 818

Y=Runs Allowed= 677

P.E. Winning Percentage: x^2/(x^2+y^2)=

(818^2)/(818^2+677^2) =**.593**

This means their WP would be expected to be .593 when it was actually .568

Pythagorean Expectation Continued

162 games in a season

(.593)*162= 96 wins

Pythagorean Record– 96-66

Actual Record– 92-70

Based on the Pythagorean Expectation, the Diamondbacks should have won 4 more games than they actually did.

Next Objective:

Is there something significant about using the exponent 2?

Which exponent is best?
Methodology

- Collect League statistics for teams from 1906-2006- runs scored, runs allowed & winning percentages
- Find Pythagorean Expectation of each team.
- Determine which exponent works best by finding the sum of squares of the residuals for each team and repeating it for each number chosen.

Which exponent is best?

Pythagorean Expectation

NL 2006	RUNS SCORED	RUNS ALLOWED	WIN PERCENTAGE	PYTHAGOREAN EXPECTATION
ARI	773	788	0.469	0.490
ATL	849	805	0.488	0.527
CHC	716	834	0.407	0.424
CIN	749	801	0.494	0.466
COL	813	812	0.469	0.501
FLA	758	772	0.481	0.491
HOU	735	719	0.506	0.511
LAD	820	751	0.543	0.544
MIL	730	833	0.463	0.434
NYM	834	731	0.599	0.566
PHI	865	812	0.525	0.532
PIT	691	597	0.414	0.573
SDP	731	679	0.543	0.537
SFG	746	790	0.472	0.471
STL	781	762	0.516	0.512
WSN	746	872	0.438	0.423
TOTAL	12337	12358		

Which exponent is best?
Pythagorean Expectation

AL 2006	RUNS SCORED	RUNS ALLOWED	WIN PERCENTAGE	PYTHAGOREAN EXPECTATION
BAL	768	899	0.432	0.422
BOS	820	825	0.531	0.497
CHW	868	794	0.556	0.544
CLE	870	782	0.481	0.553
DET	822	675	0.586	0.597
KCR	757	971	0.383	0.378
LAA	766	732	0.549	0.523
MIN	801	683	0.593	0.579
NYY	930	767	0.599	0.595
OAK	771	727	0.579	0.529
SEA	756	792	0.481	0.477
TBD	689	856	0.377	0.393
TEX	835	784	0.494	0.531
TOR	809	754	0.537	0.535
TOTALS	11262	11041		

Which exponent is best? Residual

- Residual= Pythagorean WP – Actual WP
- Example: AZ Diamondbacks
- .593-.568 = .025
- Square of residual (sum of squares)= (Pythagorean WP – Actual WP)^2
- Example: (.025)^2= .000625

Sum of Squares- exponent 2

NATIONAL	SUM SQ(2)
ARI	0.000457603
ATL	0.001488676
CHC	0.000299624
CIN	0.000756837
COL	0.000999533
FLA	9.70312E-05
HOU	2.50279E-05
LAD	6.99742E-07
MIL	0.000818759
NYM	0.001120184
PHI	4.3198E-05
PIT	0.025152281
SDP	3.8077E-05
SFG	3.87315E-07
STL	1.36028E-05
WSN	0.000237299
TOTAL	0.03154882

AMERICAN	
BAL	0.00010205
BOS	0.001158687
CHW	0.000133713
CLE	0.005201012
DET	0.000126755
KCR	2.47199E-05
LAA	0.000692466
MIN	0.000195573
NYY	1.46399E-05
OAK	0.002465403
SEA	1.80051E-05
TBD	0.000261071
TEX	0.001403978
TOR	3.4402E-06
TOTAL	0.011801513

Which exponent is best?

- Add the sum of squares for each league, each year to get the total sum of squares for exponent 2:

- **2006 NL .03154882**
- **2006 AL .011801513**
- **2001 NL .012981895**
- **2001 AL .004850473**
- **1996 NL .004758341**
- **1996 AL .005388855**
- **1991 NL .00386536**
- **1991 AL .002701488**
- **…**
- **1921 NL .001985464**
- **1921 AL .006720416**
- **1916 NL .003344076**
- **1916 AL .002204864**
- **1911 NL .014133613**
- **1911 AL .009422904**
- **1906 NL .007804765**
- **1906 AL .010944075**
- **TOTAL SS: .329895525**

Which exponent is best?

Try using 1.9 as the exponent in the formula…

x^1.9/(x^1.9+y^1.9)

Sum of Squares- exponent 1.9

NL	SUM SQ(1.9)
ARI	0.000478383
ATL	0.001388048
CHC	0.000442695
CIN	0.000667695
COL	0.000997588
FLA	0.000106251
HOU	1.98276E-05
LAD	1.80875E-06
MIL	0.000643551
NYM	0.001347683
PHI	2.49778E-05
PIT	0.024028863
SDP	6.40952E-05
SFG	6.49431E-07
STL	1.85206E-05
WSN	0.000134372
TOTAL	0.030365007

AL	SUM SQ(1.9)
BAL	3.91382E-05
BOS	0.001148364
CHW	0.000189734
CLE	0.004827567
DET	4.23841E-05
KCR	8.08756E-07
LAA	0.000753369
MIN	0.000319488
NYY	7.1868E-05
OAK	0.002612954
SEA	9.50262E-06
TBD	0.000455716
TEX	0.001288815
TOR	1.30081E-05
TOTALS	0.011772716

Which exponent is best?

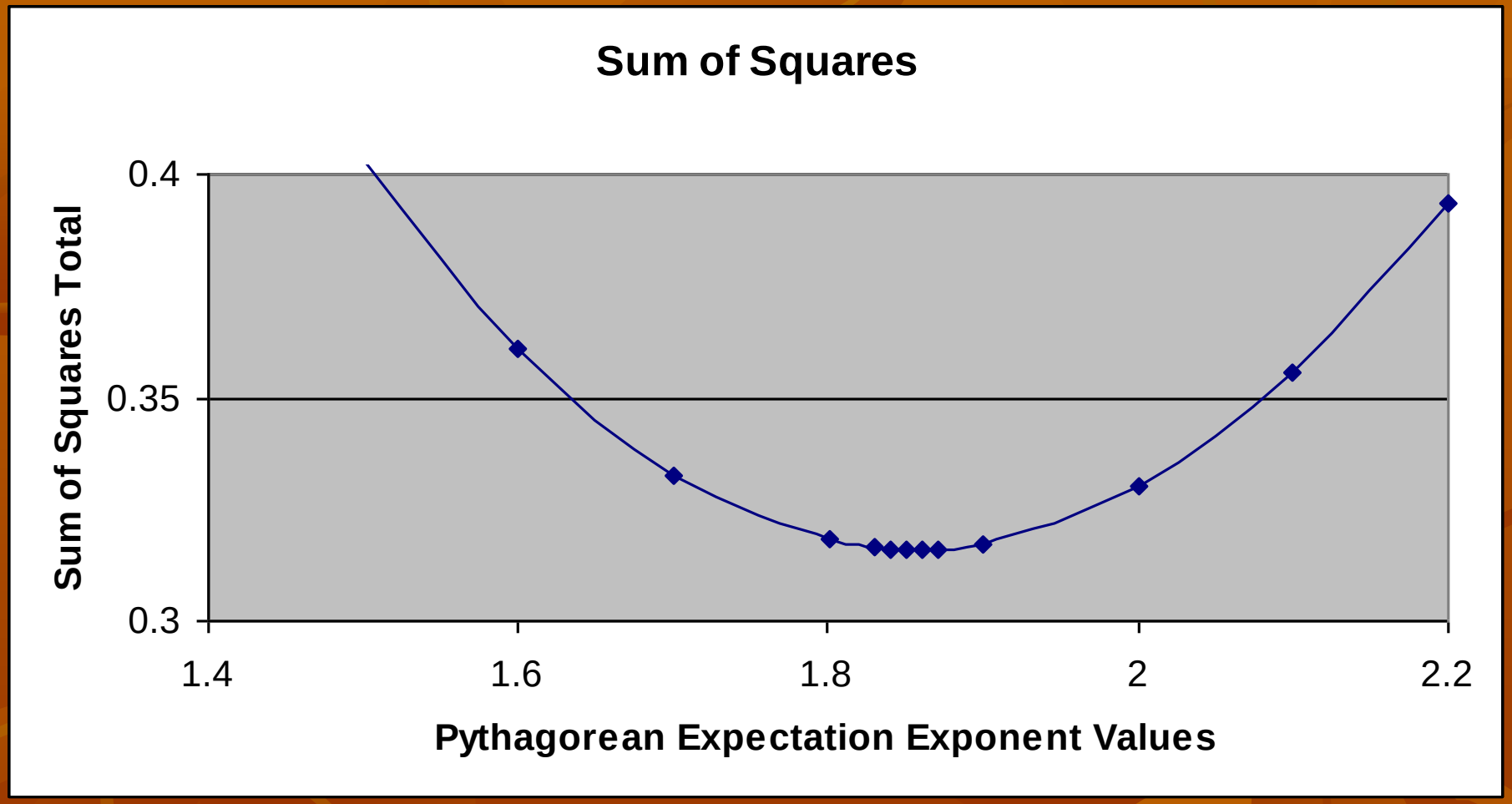
- Add the sum of squares for each league, each year to get the total sum of squares for exponent 1.9:

- 2006 NL .030365007
- 2006 AL .011772716
- 2001 NL .012521547
- 2001 AL .005637482
- 1996 NL .004973493
- 1996 AL .005096314
- 1991 NL .004029142
- 1991 AL .002883718
- …
- 1921 NL .001580867
- 1921 AL .006314575
- 1916 NL .003114606
- 1916 AL .002560682
- 1911 NL .012795209
- 1911 AL .009427654
- 1906 NL .007010311
- 1906 AL .009578747
- **TOTAL SS: .317387682**

Since 1.9 is a closer estimate for the Pythagorean Expectation formula, what about using lots of additional exponents to see if there is an even better estimate…

Sum of Squares Results

Exponent	Sum of Squares Total
1.5	0.403156
1.6	0.360806
1.7	0.332567
1.8	0.318185
1.83	0.316532
1.84	0.316251
1.85	0.316105
1.86	**0.316094**
1.87	0.316208
1.9	0.317388
2	0.329895
2.1	0.355416
2.2	0.393649

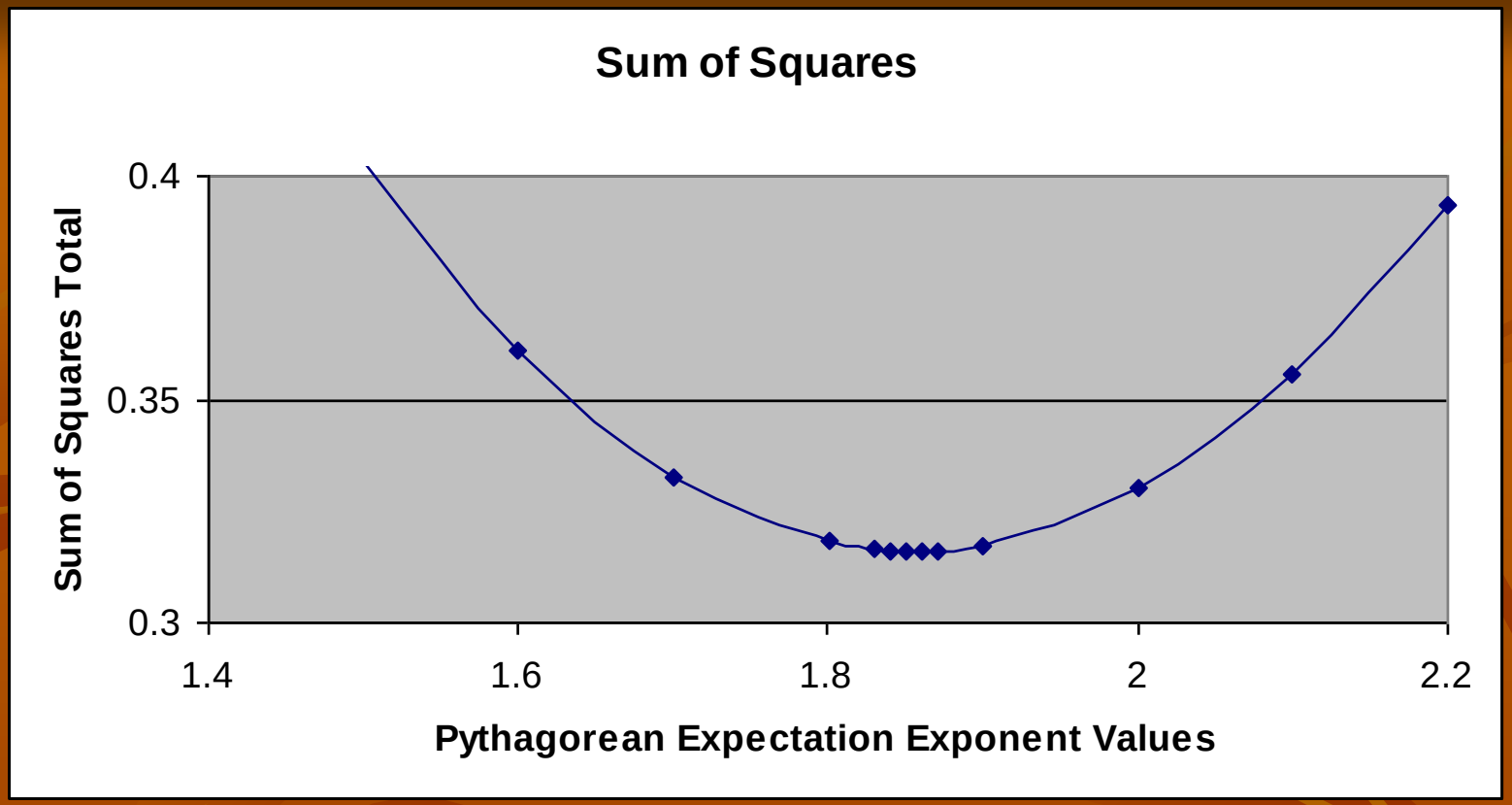

Sum of Squares
Sum of Squares Total
0.4
0.35
0.3
1.4
1.6
1.8
2
2.2
Pythagorean Expectation Exponent Values

Which exponent is best?

- Finding the quadratic regression to the points on the graph:

  y= .6717258774x^2 – 2.497560796x + 2.637625204

- Finding the vertex of the parabola, the close estimate turns out to be 1.86, the exponent chosen for the best result.



- y= .6717258774x^2 – 2.497560796x + 2.637625204

- On the calculator, this quadratic equation is the parabola that fits all these points. The parabola goes through all the points.

Conclusion

- Based on the data collected from the Sum of Squares of the years used, using an exponent of 1.86 gives a better estimate than using an exponent of 2.
- In several baseball sites using the Pythagorean Expectation, they also claim to use different exponents other than 2, without citing their reasons why.

"Empirically, this formula correlates fairly well with how baseball teams actually perform, although an exponent of 1.81 is slightly more accurate."- *Wikipedia*

"Many sabermetricians feel the results will be more accurate if, instead of SQUARING the numbers, those results are actually calculated to the power of 1.83."- *Bryan P. Douglass, FantasyBaseball.com*

Pitching Wins Championships

- “The menu for a championship in baseball is short. Good defense and good pitching.”- *Rob Parker, Detroit News*
- “It's unusual for a below-average pitching team to make it to the World Series, but you can be a below-average offensive team and make it.”- *Jeff Merron and David Schoenfield, ESPN*

Pitching Wins Championships

- The connection to the Pythagorean expectation shows why pitching wins a championship.
- Take a partial derivative of the Pythagorean expectation to find how scoring a run and allowing a run directly affects the overall winning percentage.

Partial Derivative for Pythagorean Expectation with respect to x

$$f(x, y) = \frac{x^{1.86}}{x^{1.86} + y^{1.86}}$$

$$\frac{\partial f}{\partial x} = \frac{1.86\, x^{0.86}\, y^{1.86}}{(x^{1.86} + y^{1.86})^2}$$

$$\frac{\partial f}{\partial x} = y\left(\frac{1.86\, x^{0.86}\, y^{0.86}}{(x^{1.86} + y^{1.86})^2}\right)$$

Partial Derivative for Pythagorean Expectation with respect to y

$$f(x, y) = \frac{x^{1.86}}{x^{1.86} + y^{1.86}}$$

$$\frac{\partial f}{\partial y} = -\frac{1.86\, x^{1.86}\, y^{0.86}}{(x^{1.86} + y^{1.86})^2}$$

$$\frac{\partial f}{\partial y} = -x\left(\frac{1.86\, x^{0.86}\, y^{0.86}}{(x^{1.86} + y^{1.86})^2}\right)$$

Comparison between the partial derivatives

$$\frac{\partial f}{\partial x} = y\left(\frac{1.86\,x^{0.86}\,y^{0.86}}{(x^{1.86}+y^{1.86})^2}\right)$$

$$\frac{\partial f}{\partial y} = -x\left(\frac{1.86\,x^{0.86}\,y^{0.86}}{(x^{1.86}+y^{1.86})^2}\right)$$

- Notice the expressions inside the parentheses– they are identical
  - X= runs scored Y= runs allowed
- When X>Y (for winning teams having above .500 average), the team improves their winning percentage more by allowing one fewer run than scoring one more run.
  - This may be why many say “pitching wins championships.”

Does Data from the World Series Illustrate that fact?

- To find out, calculate μ, σ of runs scored and runs allowed for all teams in the league to find a z-score for the league championship team and then compare the two z-scores found to determine if in fact the team with the better pitching z-score (runs allowed) won the World Series.

Z-Score Calculations

- First, find the mean and standard deviation of runs scored and runs allowed for each league, each year
- Next, plug those numbers into the z-score formula
- Z-Score

$$z = \frac{x - \mu}{\sigma}$$

	NL	Z-SCORE RS	Z-SCORE RA	AL	Z-SCORE RS	Z-SCORE RA
2006	***STL***	0.194733344	***-1.638839605***	DET	0.290790956	-1.412501077
2001	***ARI***	***0.780931288***	***-1.151858161***	NYY	0.205456501	-0.777743611
1996	ATL	0.204840227	-1.233741957	***NYY***	-0.012615617	-0.944884237
1991	ATL	1.564589336	-0.423192354	***MIN***	0.707827798	***-1.191288926***
1986	***NYM***	***1.996219461***	***-1.471314568***	BOS	0.779098388	-0.893670799
1981	***LAD***	***0.69635972***	-1.447698188	NYY	-0.319454682	-2.030715838
1976	***CIN***	***2.339930493***	-0.175717437	NYY	1.24462412	-1.372467953
1971	***PIT***	***1.865925943***	-0.577659308	BAL	1.543223353	-1.533702254
1966	LAD	-0.789771733	-1.906833885	***BAL***	***1.860599854***	-0.43042547
1961	CIN	0.185148906	-0.75615602	***NYY***	***1.395009492***	***-1.41882585***
1956	BRO	0.772041345	-1.047387634	***NYY***	***1.401748248***	-0.779630861
1951	NYG	0.930312174	-0.760341691	***NYY***	***1.293369515***	***-1.00588321***
1946	***STL***	1.434625642	***-1.038400753***	BOS	1.773559645	-0.562098783
1941	BRO	1.558896226	-0.971415458	***NYY***	1.184313269	***-1.321640337***
1936	NYG	0.207181637	-1.230004921	***NYY***	***1.729289013***	-1.192730924
1931	***STL***	***1.164011433***	-1.081042178	PHA	0.432341967	-1.682430111
1926	***STL***	***1.652217289***	***-0.267554315***	NYY	1.317940177	-0.149208513
1921	***PIT***	-0.145545839	***-1.078596345***	NYY	1.305893272	-0.978745281
1916	BRO	1.029260748	-0.963833913	***BOS***	-0.371233954	***-1.035533309***
1911	NYG	0.943844174	-0.965239152	***PHA***	***1.570683422***	***-1.374934551***
1906	***CHC***	1.621410355	***-1.865425992***	CHW	0.122730548	-1.158494878

How do the winners win?

- Pitching (7): ’06, ’16, ’21, ’41, ’46, ’91, ’06
- Hitting (7): ’31, ’36, ’56, ’66, ’71, ’76, ’81
- Both (6): ’11, ’26, ’51, ’61, ’86, ’01
- Neither (1): ’96

- Based on the z-scores calculated using this particular research, there is not enough evidence to conclude that pitching, in fact, wins championships.

Rounding 3rd and Heading for Home…

- The resulting data shows that the Pythagorean Expectation formula can be improved by using a slightly more complicated exponent other than 2.
- It may vary from year to year which exponent works best, but over the data used, 1.86 seems the best fit.
- Partial differentiation of the Pythagorean Expectation shows that in order to become an even better winning team, you are better off allowing one fewer run than scoring one more run.
- The term "pitching wins championships" doesn't necessarily hold true with the data collected for this project—pitching certainly helps with getting to the championship, but doesn't ensure a win.

References

- Baseball Reference. Sports Reference, Inc. 2006. October 2006. <http://www.baseball-reference.com/>.
- Douglass, Bryan P. "FB501: Advanced Sabermetrics." Fantasy Baseball. 24 Feb 2005. Nov. 2006. <http://www.fantasybaseball.com/modules/wfsection/article.php?articleid=71>.
- Merron, Jeff and David Schoenfield. "Playoff Theories: do they hold up?" ESPN.com. 29 Sept. 2005. Nov 2006. <http://sports.espn.go.com/espn/page2/story?page=baseball/one/051003>.
- "Pythagorean Expectation." Wikipedia. 26 Oct 2006. Nov 2006. <http://en.wikipedia.org/wiki/Pythagorean_expectation>.