LS88: Sports Analytics

Pythagorean Expectation

Pythagorean Expectation

➔ From Bill James (circa 1980)

$$\text{Pythagorean Expectation Win Pct} = \frac{\text{Runs Scored}^2}{\text{Runs Scored}^2 + \text{Runs Allowed}^2}$$
$$= \frac{(\text{Runs Scored / Runs Allowed})^2}{1 + (\text{Runs Scored / Runs Allowed})^2}$$

Pythagorean Expectation

➔ An alternate version from Bill James

$$\text{Pythagorean Expectation Win Pct} = \frac{\text{Runs Scored}^{1.83}}{\text{Runs Scored}^{1.83} + \text{Runs Allowed}^{1.83}}$$

Pythagorean Expectation for NBA

➔ Applied to the NBA (Daryl Morey circa 1994)

$$\text{NBA Pythagorean Expectation Win Pct} = \frac{\text{Points For}^{13.91}}{\text{Points For}^{13.91} + \text{Points Againts}^{13.91}}$$

Pythagorean Expectation for NFL

➔ From *Football Outsiders Almanac*

$$\text{NFL Pythagorean Expectation Win Pct} = \frac{\text{Runs Scored}^{2.37}}{\text{Runs Scored}^{2.37} + \text{Runs Allowed}^{2.37}}$$

Examples

Examples from right now:

- ➔ Boston Red Sox, currently 97 wins and 46 losses
    - Runs: 5.4
    - Runs Allowed: 3.9
    - Expected wins: 92 wins
- ➔ Houston Astros, currently 89 wins and 53 losses
    - Runs: 5.0
    - Runs Allowed: 3.31
    - Expected wins: 97 wins

Examples

Some other examples:

Team	W	L	R	RA	pythW	pythL	pythLuck
Sea	78	64	4.1	4.5	65	77	13
Oak	86	57	4.8	4.2	80	63	6
Col	78	63	4.7	4.7	70	71	8
LAD	77	65	4.7	3.8	85	57	-8

Pythagorean Expectation

We know you need to score more than your opponent

So what’s the premise of this model?

- ➔ Give me a team’s run scored/allowed profile, and I can give you an expected winning percentage
- ➔ We can use that to quantify a measure of “luck” a team has had
  - ◆ Games are far too intricate to just say it’s *just* luck, but it is a quick description of under/over-performance

Pythagorean Expectation

But it's much more than just expected number of wins from run scoring!

We're trying to win games (duh!)

But a lot happens in a game where at the end you only get a simplified summary: a W or a L

- Known as a binary outcome

The model tells us how run scoring drives winning and losing

- Or point scoring in other sports

Quantifying a Player's Value

Take our measurables (all that data collected in box scores)

…to get a value in runs from our models to come

…and then get a value in wins from Pythagorean Expectation

- ➔ Value in *runs* (from FanGraphs):
  - ◆ Mike Trout has been worth 63.8 runs above an average player on offense
  - ◆ Mookie Betts has been worth 59 runs above an average player on offense
- ➔ Value in *wins*
  - ◆ Trout and Betts have been worth about 6 wins above an average player

Quantifying a Player's Value

Key Takeway!

If we can measure a player's performance in terms of runs, we can measure their value in terms of wins

Without the model, we'd need to directly measure a player's effect on winning

THIS IS HARD TO DO!

Also, sometimes you can do everything right up until the end and still get a loss

Pythagorean Expectation

Some comments:

- ➔ This is the premise of Wins Above Replacement (WAR) which we'll cover
  Measure performance in terms of runs and convert to Wins using a model like Pythagorean Expectation

- ➔ Okay, so what drives scoring? We'll get to that…

Demo

We need to make more sense of this by getting our hands on the data.

Load the demo

“Pythagorean Expectation - The Relationship between Runs and Wins”

Deriving Pythagorean Expectation

- Team Quality (run ratio)
- Win Pct proportional to team quality
- Odds ratio, log odds ~ log quality

Deriving Pythagorean Expectation

➔ We know what the formula is, but what's the motivation?

$$\begin{aligned}\text{Pythagorean Expectation Win Pct} &= \frac{\text{Runs Scored}^2}{\text{Runs Scored}^2 + \text{Runs Allowed}^2} \\ &= \frac{(\text{Runs Scored / Runs Allowed})^2}{1 + (\text{Runs Scored / Runs Allowed})^2}\end{aligned}$$

Odds and Log-Odds

- ➔ A team's winning percentage is its likelihood of winning
- ➔ The odds of winning is given by

$$\text{Odds} = \frac{\text{Win Pct}}{1 - \text{Win Pct}} = \frac{\text{Win Pct}}{\text{Loss Pct}}$$

Odds and Log-Odds

- ➔ A team's winning percentage is its likelihood of winning
- ➔ The odds of winning is given by

$$\text{Odds} = \frac{\text{Win Pct}}{1 - \text{Win Pct}} = \frac{\text{Win Pct}}{\text{Loss Pct}}$$

Example: A team wins 60% of its games, its odds of winning is 3/2

Odds and Log-Odds

- ➔ Take a team with 3/2 odds of winning (.600 winning percentage)
- ➔ If we double the odds to 3/1, the team will win at a .750 winning percentage (the team will win 3 games for every 1 game it wins)

Adjusting the odds adjusts how many times we expect an outcome against its opposite

Because odds are a ratio, they work best when being multiplied

But we like sums insteads ratios/multiplications...

Odds and Log-Odds

- In statistics/mathematics, it's better to deal with sums than ratios
- Hence we consider the log-Odds:

$$\log \text{Odds} = \log \frac{\text{Win Pct}}{\text{Loss Pct}}$$

If we increase the log-Odds by log 2, we double the odds

$$\exp(\log \text{Odds} + \log 2) = 2 \times \text{Odds}$$

Odds and Log-Odds

If we increase the log-Odds by log 2, we double the odds

$$\exp(\log \mathrm{Odds} + \log 2) = 2 \times \mathrm{Odds}$$

If we double the log-Odds by log 2, we *square* the odds

$$\exp(2 \log \mathrm{Odds}) = \mathrm{Odds}^2$$

Team Quality

➔ We're going to define team quality through the Run Ratio (RR)

$$\text{Run Ratio} = \frac{\text{Runs Scored}}{\text{Runs Allowed}}$$

We could use Run Differential (RD) but this is better:

RR better adapts to varying levels of run scoring.

RD does not

➔ The opponent quality is defined as 1 / RR

Derivation 1: Team Quality

➔ We posit winning percentage is proportional (roughly) to team quality

$$\text{Win Pct} = \frac{\text{Run Ratio}}{\text{Run Ratio} + 1/\text{Run Ratio}}$$

➔ Run Ratio is a fraction, if we multiply through to clear the fraction, we get:

$$\text{Win Pct} = \frac{\text{Runs Scored}^2}{\text{Runs Scored}^2 + \text{Runs Allowed}^2}$$

Pythagorean Expectation comes from modeling team quality!

Derivation 2: Log-Odds

➔ We posit log-Odds is twice the log Run Ratio:

$$\log \text{Odds} = 2 \times \log \text{Run Ratio}$$

$$\Rightarrow \text{Odds} = \text{Run Ratio}^2$$

$$\Rightarrow \text{Win Pct} = \frac{\text{Runs Scored}^2}{\text{Runs Scored}^2 + \text{Runs Allowed}^2}$$

Once again, the algebra works out to provide Pythagorean Expectation

Computing the Exponent Empirically

Okay, we posited this relationship:

$$\log \mathrm{Odds} = 2 \times \log \mathrm{Run\ Ratio}$$

Can we empirically verify this is valid (or close)?

Also, the coefficient 2 turns into the exponent in the formulas. Can we determine a different coefficient/exponent?

Yes!

Continuing the Demo...

Runs Per Win

Compute the change in Pythagorean Wins per change in Run Ratio:

$$\begin{aligned}\text{Change in Wins per Run Ratio} &= \frac{\text{Change in Wins}}{\text{Change in Runs Ratio}} \\ &= \frac{d}{d\text{Runs Ratio}}\ \text{Pythagorean Expectated Wins}\end{aligned}$$

Runs Per Win

Differentiation and algebra gives us:

$$\begin{aligned}
\text{Change in Wins per Run Ratio} &= \frac{d}{d\text{Run Ratio}}\text{Games} \times \left(\frac{\text{Run Ratio}^2}{1+\text{Run Ratio}^2}\right) \\
&= \frac{\text{Games}}{\text{Run Ratio}^3}\left(\frac{\text{Run Ratio}}{1+\text{Run Ratio}^2}\right)^2 \\
&= \frac{\text{Games}}{\text{Run Ratio}^3}(\text{Pythagorean Expected Win Pct})^2
\end{aligned}$$

Runs Per Win

Some more algebra gives us:

$$\text{Change in Runs Ratio per Win} = \frac{1}{2 \times \text{Pythagorean Wins}} \times \frac{\text{Run Ratio}^3}{\text{Pythagorean Win Pct}}$$

Continuing the Demo...

Runs Per Win

- ➔ Run Ratio is Runs Scored over Runs Allowed
- ➔ If we multiply out by Runs Allowed we get what we want

$$\text{Change in Runs Scored per Win} = \\ \text{Games} \times \text{Runs Allowed per Game} \times \text{Change in Runs Ratio per Win}$$