

Is It Harder To Get a “W” In A Pitchers Park? (And a Mystery Uncovered)

By Gary Fletcher

I had this idea about Sandy Koufax. I thought that the magnificent Won / Lost records of his peak years might have been even better in a different home park. My reasoning went something like this:

1. Pitchers Parks reduce scoring
2. Less runs should equal closer games
3. Games decided by many runs should favour good teams much more than games decided by just a run or two, so...
4. There should be a wider gap in W/L percentage when the margin of victory is larger

Or, try this. Imagine a .600 team. Could such a team be .650 in blowouts, but just .550 in one run games? FMA numbers, to be sure, but you get the idea.

How about: The closer the score the closer we get to coin flip territory. Dodger Stadium was an extreme pitchers park and certainly helped Koufax' supporting statistics, but may have hindered his W/L record. Of course when you are 25 and 5 in 40 starts there isn't much room for improvement. Well, let's find out if there is any supporting evidence.

So, (thanks to Retrosheet) I am able to count how many times X runs are scored in a game, and also how often a game was decided by a specific score.

I built a database of all scores from 1901 through 2014 (that's 186,377 games). I started with 1901 because (due to another article (See "Just How Tough Is It?") I believe I have shown that the competitive environment of MLB reached a stable level around that time, enhanced by the stability of the two leagues with 8 franchises in each league.

A few fun notes:

- Starting with the 1901 season there have been 786 games that ended in a tie, but the last tie was in 2005. The tie is extinct to the same extent that smallpox is extinct
- The highest scoring game in this period saw 49 runs. Cubs beat the Phillies 26-23 on August 25, 1922. You have probably heard of this game.
- The biggest margin of victory in this time period? 30-3, or 27 runs. Texas over Baltimore August 27, 2007. You have probably heard of this one, too.
- The most common score in MLB history is 3-2. There were 10,718 games that ended 3-2.
- Second most common score is 4-3, occurring 10357 times.

- There are 35 scores that are tied for least common, with one occurrence each. The lowest scoring being a 20-0 game, the highest scoring being mentioned above. Oh, well, I'll run a table here:

W Score	L Score		W Score	L Score		W Score	L Score		W Score	L Score
26	23		24	11		25	8		26	5
23	22		29	6		29	4		23	7
25	13		18	16		30	3		27	3
20	16		24	10		19	13		24	5
22	14		26	8		22	10		23	5
23	13		28	6		23	9		22	5
24	12		18	15		25	7		25	2
19	16		19	14		18	13		22	3
22	13		22	11		23	8		23	1
23	12		24	9		25	6		20	0

So, if you see one of these you can tell people next to you that this is only the second time in the last 115 years this score has been posted.

One more thing. The following scores – all certainly within the realm of possibility – have not occurred in MLB 1901 through 1914:

22-0 24-3 25-5 21-9 18-13

Okay back to *this* article. I'll run a table (from the Home Team viewpoint):

Victory Margins	Wins	Losses	W %	Distance from .500	Victory Margins	Wins	Losses	W %	Distance from .500
1	34424	21839	0.612	0.112	15	110	90	0.550	0.050
2	16958	17000	0.499	-0.001	16	60	66	0.476	-0.024
3	13680	13246	0.508	0.008	17	34	32	0.515	0.015
4	10587	10009	0.514	0.014	18	24	30	0.444	-0.056
5	7754	7308	0.515	0.015	19	16	13	0.552	0.052
6	5512	5137	0.518	0.018	20	5	6	0.455	-0.045
7	3751	3657	0.506	0.006	21	6	6	0.500	0.000
8	2744	2495	0.524	0.024	22	3	5	0.375	-0.125
9	1851	1578	0.540	0.040	23	1	3	0.250	-0.250
10	1250	1041	0.546	0.046	24	1	0	1.000	0.500
11	724	656	0.525	0.025	25	1	0	1.000	0.500
12	508	440	0.536	0.036	26	0	0		
13	293	280	0.511	0.011	27	0	1	0.000	-0.500
14	170	186	0.478	-0.022					

In a rough way this kinda sorta supports one part of the theory: that the closer the games the closer the results will be to .500. At least until we get to games decided by 11 or more runs when the results either start to go backwards or get rather volatile (a small sample size phenomenon).

But the really striking thing is that games decided by 1 run are so dramatically far away from .500 (.612 W% from the home team viewpoint) compared to games decided by 2 runs (.499 W% from the home team viewpoint). After 2 runs there is a fairly regular movement away from .500...to .508, then .514, .515, .518...well, you can see the table.

You might be wondering if using such a long time period is the wrong way to go. But I checked shorter time periods and the results hold true to form no matter what you do.

Here, let's look at the last 10 years (2005 – 2014):

Victory Margins	Wins	Losses	W %	Distance from .500	Victory Margins	Wins	Losses	W %	Distance from .500
1	2015	1339	0.601	0.101	15	5	6	0.455	-0.045
2	1040	1028	0.503	0.003	16	6	3	0.667	0.167
3	831	761	0.522	0.022	17	3	0	1.000	0.500
4	688	657	0.512	0.012	18	2	3	0.400	-0.100
5	490	501	0.494	-0.006	19	0	1	0.000	-0.500
6	362	349	0.509	0.009	20	0	0		
7	243	255	0.488	-0.012	21	0	0		
8	164	159	0.508	0.008	22	0	0		
9	114	114	0.500	0.000	23	0	0		
10	73	62	0.541	0.041	24	0	0		
11	51	48	0.515	0.015	25	0	0		
12	31	35	0.470	-0.030	26	0	0		
13	20	22	0.476	-0.024	27	0	1	0	-0.5
14	8	14	0.364	-0.136					

There are differences. Of course there are. But again, what is it with one run games, why do they so dominate the question of who wins?

Before I attempt to answer that, let's dispose of the Sandy Koufax question. I think that the answer is yes, lower scoring games tend to depress W/L records a little bit. But when combined with the dramatic counter tendency (that one run games move dramatically away from .500), the overall effect seems negligible. I think Sandy earned pretty much everything he got from Won / Lost records.

You may be questioning why I derived W/L records from Home Teams vs Road Teams. Well, you have to split them up in some way that is fair and equal or else everything just

devolves to .500. I chose home vs road, could have been from the home team viewpoint or the road team viewpoint, flip a coin.

Having done that, am I just measuring the home field advantage? To some extent, yes, but why are one run games favouring the home team by .600 to .500? Indeed, the one run games beat the two, three, four, five, six, seven, eight, nine and ten run margin games combined by .612 to .510. Why is that?

I don't know. But having started and disposed with the simple question I now find myself faced with a more intriguing phenomenon. Maybe this is something everyone else already knows...but I didn't, and I don't.

I thought I would look at the W/L records for specific scores. After all, a one run game is typically 1-0 or 3-2 or 6-5, stuff like that, but it can also be 11-10 or 16-15.

Here is the one run margin table (using 2005-2014 data, just through the 10 to 9 score):

One Run Margins					
Score		Wins	Losses	W %	Distance from .500
1	0	120	77	0.609	0.109
2	1	277	201	0.579	0.079
3	2	365	232	0.611	0.111
4	3	427	280	0.604	0.104
5	4	320	194	0.623	0.123
6	5	215	155	0.581	0.081
7	6	140	91	0.606	0.106
8	7	71	50	0.587	0.087
9	8	37	34	0.521	0.021
10	9	18	14	0.563	0.063

I had thought that perhaps the one run category would be dominated by one particular score (1-0 perhaps) but not so. The home teams dominate one run games throughout.

Let's take a look at two run margins. Remember that this was the one category in which home teams were sub .500, albeit at .499.

Two Run Margins					
Score		Wins	Losses	W %	Distance from .500
2	0	90	98	0.479	-0.021
3	1	176	160	0.524	0.024
4	2	214	203	0.513	0.013
5	3	180	175	0.507	0.007
6	4	167	150	0.527	0.027
7	5	94	95	0.497	-0.003
8	6	48	70	0.407	-0.093
9	7	31	37	0.456	-0.044
10	8	25	22	0.532	0.032
11	9	10	8	0.556	0.056

Again, pretty consistent among the various 2 run margin scores.

Three Run Margins					
Score		Wins	Losses	W %	Distance from .500
3	0	101	98	0.508	0.008
4	1	154	126	0.550	0.050
5	2	179	140	0.561	0.061
6	3	143	154	0.481	-0.019
7	4	109	106	0.507	0.007
8	5	63	55	0.534	0.034
9	6	38	45	0.458	-0.042
10	7	17	22	0.436	-0.064
11	8	12	10	0.545	0.045
12	9	12	3	0.800	0.300

We are seeing a bit more deviation score to score now. Will that continue as we look at higher run margins?

Four Run Margins					
Score		Wins	Losses	W %	Distance from .500
4	0	91	88	0.508	0.008
5	1	151	117	0.563	0.063
6	2	143	150	0.488	-0.012
7	3	119	112	0.515	0.015
8	4	80	79	0.503	0.003
9	5	48	49	0.495	-0.005
10	6	26	31	0.456	-0.044
11	7	13	12	0.520	0.020
12	8	9	9	0.500	0.000
13	9	3	8	0.273	-0.227

Five Run Margins					
Score		Wins	Losses	W %	Distance from .500
5	0	89	66	0.574	0.074
6	1	106	111	0.488	-0.012
7	2	103	109	0.486	-0.014
8	3	84	76	0.525	0.025
9	4	47	61	0.435	-0.065
10	5	21	33	0.389	-0.111
11	6	25	21	0.543	0.043
12	7	10	10	0.500	0.000
13	8	2	6	0.250	-0.250
14	9	2	4	0.333	-0.167

It looks like the answer is “yes.” As more runs are scored the variance of W / L % between scores does seem to increase.

Six Run Margins					
Score		Wins	Losses	W %	Distance from .500
6	0	69	39	0.639	0.139
7	1	88	79	0.527	0.027
8	2	72	71	0.503	0.003
9	3	55	58	0.487	-0.013
10	4	30	44	0.405	-0.095
11	5	18	25	0.419	-0.081
12	6	15	17	0.469	-0.031
13	7	5	7	0.417	-0.083
14	8	4	4	0.500	0.000
15	9	3	2	0.600	0.100

Of course we are now seeing the effects of small sample sizes as higher total runs are being scored.

Seven Run Margins					
Score		Wins	Losses	W %	Distance from .500
7	0	45	37	0.549	0.049
8	1	63	59	0.516	0.016
9	2	43	51	0.457	-0.043
10	3	39	46	0.459	-0.041
11	4	20	28	0.417	-0.083
12	5	21	18	0.538	0.038
13	6	6	6	0.500	0.000
14	7	5	3	0.625	0.125
15	8	1	5	0.167	-0.333
16	9	0	2	0.000	-0.500

Eight Run Margins					
Score		Wins	Losses	W %	Distance from .500
8	0	29	30	0.492	-0.008
9	1	43	29	0.597	0.097
10	2	28	30	0.483	-0.017
11	3	28	32	0.467	-0.033
12	4	19	18	0.514	0.014
13	5	9	8	0.529	0.029
14	6	5	6	0.455	-0.045
15	7	1	2	0.333	-0.167
16	8	1	2	0.333	-0.167
17	9	1	0	1.000	0.500

Trust me, the variance gets more and more extreme as the victory margins get higher and, consequently and naturally, the total runs scored have to be higher (you can't win by 8 runs without the total runs being 8 or more, after all).

Bill James once stated (paraphrasing now) that "...the closer the game the better the chance that the best team will lose." That assertion isn't challenged here, not exactly. This is an examination of all teams regardless of quality.

So, what is being shown here?

What I think is being shown here is what the Home Team advantage is by margin of victory. And oddly enough, the lower the margin of victory the higher the winning percentage is for the home team.

GFletch – September 1, 2015

Postscript December 3, 2015

I decided to ask Bill James about this...he responded:

Well, there's an obvious bias in the data. If they're ahead by one run after 8 1/2 innings, the home team doesn't bat in the ninth, so the margin stays at one run. If they're tied after 8 1/2

innings, the home team bats until they score one run, and then the game ends as soon as the home team scores one run, so there's a one-run victory again. This creates a cluster of one-run victories in home games, but it doesn't really give us any deeper insight into the game.

I asked him if I could quote him here, he graciously said “Sure” and so...now I can sleep at night.