A Modification of OPS: Widely Used to Measure a Baseball Batter's Performance

Chulmin Kim[1]

Abstract

Batting average (BA), home runs (HRs), and runs batted in (RBIs) have been the most dominant statistics to measure a baseball batter's performance. Since each of those contains a meaningful interpretation but also some drawback to explain a batter's ability at the same time, often we use those three together. Slugging percentage (SLG) and on-base percentage (OBP) have been used as alternatives of the traditional three statistics. SLG measures how often a batter hits and how valuable the hits are and OBP measures how often a batter reaches bases. Whereas SLG ignores reaching bases by hits by pitched ball or walks, OBP is limited to measure the quality of the hits. A combination of these two is called OPS, the sum of OBP and SLG, which has become more widely used. We introduce a variation of OPS, WOA (weighted offensive average), which is a single number explaining not only a batter's hitting performance but also his non-hitting performance to generate runs for his team such as stolen bases, walks, and etc. This newly developed statistic is based on major league team statistics from the year 2000 to the year 2008.

Key Words: regression, correlation coefficient, longitudinal data, baseball statistics

1. Introduction

1.1 Who is more valuable batter? Ichiro? Or Matsui?

Two best Japanese batters to play in the major league baseball may be Ichiro Suzuki of New York Yankees (2012~) and Hideki Matsui of Tampa Bay Devil Rays (2012~). They both similarly began their professional baseball career in Japan and moved to the major league after spending 9 years in Japan. As shown in **Table 1**, Ichiro hits .353 BA with 118 HR and 529 RBIs for Orix Buffalos from 1992 to 2000 and Matsui hits .304 BA with 332 HR and 889 RBIs for Yomiuri Giants from 1994 to 2002. As we can see from those descriptive statistics Ichiro is more likely a slap hitter and Matsui is more likely a slugger. Even though they have different batting style, they have been very valuable for their teams. Who is more valuable?

Table 1: Descriptive career statistics for Ichiro and Matsui in Japan and in major league.

	Japanese league				Major league			
Player	Hits	BA	HR	RBIs	Hits	BA	HR	RBIs
Ichiro	1278	.353	118	529	2597	.322	104	656
Matsui	1390	.304	332	889	1253	.282	175	760

[1] School of Mathematical Sciences, Rochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, NY 14623, U.S.A. Email: cxksma@rit.edu

We conduct a comparison between Ichiro and Matsui in 2004. As shown in **Table 2**, Ichiro and Matsui made about the same salary in the year. Ichiro performed better than Matsui in the categories of hits, BA, and OBP in 2004. On the other hand, Matsui did better in HR, RBIs, SLG, and OPS in the year. Especially Matsui has higher OPS than Ichiro (.912 versus .869). However Ichiro made higher salary in 2008 ($17 mil. versus $13 mil.) Does it mean that OPS is not a big factor to explain a batter's salary? Or any better single offensive statistic can explain their difference in salary more precisely?

Table 2: Descriptive baseball statistics for Ichiro and Matsui in the major league in 2004 and salary in 2004 and 2008. Note: The player with **bold** is better in the category in 2004.

Player	Salary (2008)	Salary (2004)	Hits	BA	HR	RBIs	OBP	SLG	OPS
Ichiro	**$17 mil.**	$6.5 mil.	**262**	**0.372**	8	60	**0.414**	0.455	0.869
Matsui	$13 mil.	**$7 mil.**	174	0.298	**31**	**103**	0.390	**0.522**	**0.912**

1.2 Why new single offensive statistic needed?

James (1985) explains why he believes runs created (RC) is an essential measurement of batting ability in his book. "*With regard to an offensive player, the first key question is how many runs have resulted from what he has done with the bat and on the base-paths. Willie McCovey hit .270 in his career, with 353 doubles, 46 triples, 521 home runs and 1,345 walks - but his job was not to hit doubles, nor to hit singles, nor to hit triples, nor to draw walks or even hit home runs, but rather to put runs on the scoreboard. How many runs resulted from all of these things?*" After that many researchers have focused on making a model to measure a batter's ability to contribute to generate runs for his team. Another effort to measure a batter's performance more precisely was introduced by by Thorn and Palmer (1984). It is called an On-base plus slugging (OPS) which is simply the sum of OBP and SLG. There have been studied by several slightly varied statistics of OPS. One example of those is called Gross production average (GPA) introduced by Gleeman (2003). GPA is obtained by (the sum of SLG and 1.8 times OBP)/4. It presents better relative weight to its two components OBP and SLG and its scale is somewhat similar to the already familiar BA. However both OBP and SLG contain the interpretation of BA, any statistics based on those two would depend on BA doubly. We introduce a new variation of OPS, WOA (weighted offensive average), which solves the problem of the redundancy of BA and explains a batter's non-hitting performance to generate runs for his team such as stolen bases, walks, and etc. Basic baseball terminologies are displayed in **Table 3**.

2. Dataset and models

2.1 Dataset

As we mentioned before, we only consider offensive statistics not pitching statistics. The dataset described in this paper contains team's batting statistics as well as players' batting statistics for 14 American League (AL) and 16 National League (NL) from 2000 to 2008. Therefore we use 270 teams' offensive statistics (30 teams for 9 years). And we use all 147 players' and 142 players' batting statistics those who were qualified in the seasons of 2008 and 2012, respectively.

Table 3: Basic baseball terminologies and their abbreviations

Abbreviation	Meaning	Abbreviation	Meaning
AB	At Bats	BA	Batting Average
PA	Plate Appearances	BB	Base on Balls
R	Runs	HBP	Hit By Pitched Ball
H	Hits	BB	Base on Balls
2B	Doubles	SB	Stolen Bases
3B	Triples	CS	Caught Stealing
HR	Home Runs	SF	Sacrifice Flies
TB	Total Bases	SLG	Slugging Percentage
RBI	Runs Batted In	OBP	On Base Percentage

2.2 Models

Batting average measures the percentage of hits a batter earns for his total at bats. It provides a strong measure of a batter's ability to produce hits. However it fails to measure the quality of hits and also fails to detect a batter's ability to reach bases by non-hits such as BB and HBP. As the formula shown in **Table 4**, slugging measures the quality of a batter's hits in addition to the ability to produce hits. However still it cannot detect a batter's ability to reach bases by non-hits. Let us assume that there are two players A and B. And we assume that player A has AB=5 and H=1 (1 HR) and player B has AB=5 and H=4 (4 singles). Then the SLG for player A and player B are the same as .800 whereas the BA for player A=.200 is much lower than the one for player B=.800. Even though their SLG are same, player B is not a power hitter like player A though his ability producing hits is good. An alternative to solely measure a batter's ability as a power hitter is called the Isolated power (ISO). By the formula given in **Table 4**, ISO for player A=.600 (.800-.200) is much higher than the one for player B=.000=(.800-.800). We introduce a variation to ISO which is called the Pure slugging percentage (pSLG). pSLG is defined by ISO/(3*BA)=(TB-H)/(3*H). Note that ISO measures how many extra bases per AB but pSLG measures how many extra bases per hit over three. It means the theoretical scale for pSLG is between 0 and 1. On-base percentage (OBP) accounts for a batter's ability to reach bases not only by hits but also by non-hits such as BB or HBP. However it fails to measure the quality of hits and considers BB and HR as the same value.

Table 4: Baseball statistics and their formulas

Statistic	Formula	Statistic	Formula
PA	AB+BB+HBP+SF	SLG	TB/AB
BA	H/AB	OPS	OBP+SLG
TB	H+2B+2*3B+3*HR	GPA	(1.8*OBP+SLG)/4
OBP	(H+BB+HBP)/PA	ISO	SLG-BA=(TB-H)/AB

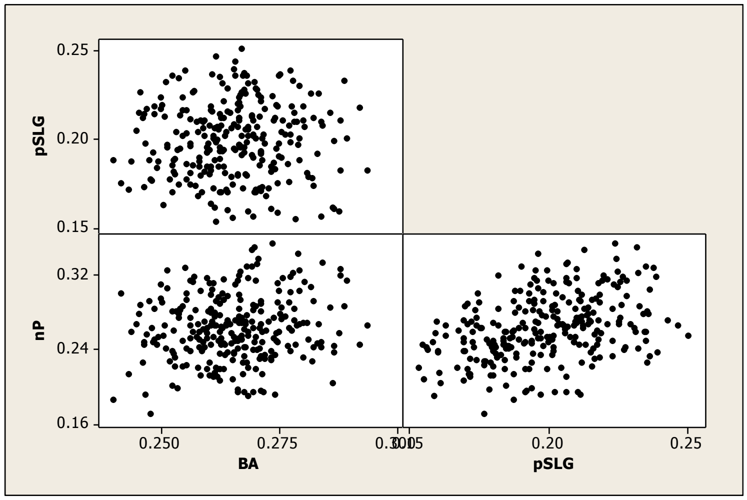
As we see in **Table 5**, in response to the deficiencies of SLG and OBP, many have turned to combinations of OBP and SLG. RC (Runs Created) and TA (Total Average) are examples to use a mixture of OBP and SLG). On-base plus slugging (OPS) has been more popular because of its easy formula as a short form to measure contribution as a batter. Let us consider another example. Let us assume that Player C has PA=20, AB=17, H=3 (3 HR), and BB=3 and player D has PA=20, AB=15, H=6 (5 singles and one 2B), and BB=5. Then (BA, OBP, SLG) for player C=(.176, .300, .706) and (BA, OBP, SLG) for player D=(.400, .550, .467). Thus player C is much better than player D in BA and OBP but not in SLG. It turns out the OPS for player C (1.006) is almost the same as the one for player D (1.017) even we may feel player D is better because he generates fewer outs. Since SLG and OPS are highly correlated (r=.97), while the correlation coefficient between OBP and OPS is .89, OPS may not differentiate SLG. And it's harder to generate .100 of OBP than .100 of SLG. GPA (Gross Production Average) is basically the weighted average of OBP and SLG with the weights 1.8 and 1. The weights come from the linear weights in the regression equation of RPG (Runs per game for a team) on a team's OBP and SLG. And GPA makes a balance of OBP and SLG in terms of their correlation coefficients with GPA (.94 versus .93). From our dataset, the regression results: RPG = -5.94 + 10.7*SLG + 18.5*OBP with R^2=90.4%. The ratio of 18.5 and 10.7 is 1.73 which is close to 1.8 of the weight in GPA.

Table 5: The comparison of several batting statistics. The newly developed statistic WOA with **bold** is able to measure all of these categories and its scale is similar to BA.

Statistics	Accuracy	Power	Reaching Bases by non-hits	Running	Q1	Q3
BA	O	X	X	X	.261	.297
SLG	O	O	X	X	.406	.506
ISO	X	O	X	X	.134	.226
pSLG	X	O	X	X	.162	.267
OBP	O	X	O	X	.328	.373
OPS	O	O	O	X	.743	.874
GPA	O	O	O	X	.253	.292
WOA	**O**	**O**	**O**	**O**	**.263**	**.298**

Since our explanatory variables SLG and OBP are highly correlated (r=.75), there exists "*multicollinearity*". Kutner (2004) says in his book: "*The simple interpretation of regression coefficients is often unwarranted with highly correlated explanatory variables*". Since BA is a component of both OBP and SLG, we may want break OBP and SLG down into BA and some non-BA part. And hopefully they are not highly-correlated to avoid the effects of "*multicollinearity*". Let us define reaching bases by non-hitting performace (nP) measures a batter's ability to reach the bases by non-hits such as BB, HBP, or SB. It is obtained by 2.8*(B%-CS%) + SB%, where B%=(BB+HBP)/PA, CS%=CS/PA and SB%=SB/PA. Here the weights 2.8 come from the linear weights in the regression equation of RPG on a team's B%, CS%, and SB%. Then we realize OBP=(H+BB+HBP)/PA=(H/PA)+(BB+HBP)/PA≈BA+B% and SLG=BA+ISO to use the formulas given in **Table 4**. Since OBP≈BA+B% and SLG=BA+ISO, we may consider the regression of RPG on BA, ISO, and B%. Since pSLG gives a similar interpretation with ISO (r=.94) and nP gives a similar interpretation with B% (r=.92). pSLG and nP would replace by ISO and B%. Another advantage to use them as explanatory variables along with BA to explain RPG is because they have much smaller correlation with BA (r=.02 for pSLG and r=.10 for nP) as shown in **Figure 1**. And also nP accounts for a batter's running ability. Now we regress RPG on BA, pSLG and WBS%. From 2000 to 2008 (14 American and 16 National league teams for nine seasons=270 teams), the regression result is that RPG = -7.14 + 34.3*BA + 8.29*pSLG + 4.31*nP with R^2=90.9%. The ratios of 34.3, 8.29, and 4.31 are similar to 8:2:1 which is used in the new proposed statistic WOA (weighted offensive average). The formula for WOA is given by WOA = (8*BA + 2*pSLG + nP)/10.5. The reason why we divide by 10.5 is to make the scale of WOA similar to already familiar BA. Then the new regression is given by RPG = -7.12 + 44.8*WOA with R^2=90.9%.

Figure 1: The matrix plot of BA, pSLG, and nP.



As shown in **Table 6**, regression coefficients of RPG on WOA by year and by league have been consistent.

Table 6: Regression coefficients of RPG on WOA by year and by league

	intercept	slope
TOT	-7.12	44.8
AL	-6.62	43.1
NL	-7.15	44.7
2000	-7.15	45.1
2001	-6.54	42.7
2002	-7.30	45.5
2003	-6.82	43.7
2004	-8.13	48.5
2005	-6.95	43.9
2006	-7.20	44.9
2007	-6.65	43.0
2008	-6.29	41.6

And as shown in **Figure 2**, WOA explains RPG as the best single statistic over OPS and GPA. The correlation coefficients between RPG and OPS, GPA and WOA are .946, .951, and .953, respectively. **Table 7** shows the correlation comparison between RPG and pSLG, nP, ISO, BA, OBP, SLG, OPS, GPA, and WOA by year and by league. As always WOA explains RPG at the best even though the differences are slight. As shown in **Figure 3**, we are confident that WOA is a good predictor for generating runs for a team through the comparative box-plots of RPG and FRPG (fitted RPG on WOA). The descriptive statistics for BA and WOA are almost identical. As shown in **Table 8**, (minimum, Q_1, median, Q_3, maximum) for BA and WOA are (.240, .259, .266, .272, .294) and (.239, .258, .266, .272, .292), respectively. And (mean, SE) for BA and WOA are (.266, .0006), respectively. **Table 9** shows the top ten WOA players in 2004. There exist some agreements between WOA and OPS. However some disagreement also exists. Let us go back to the example of Ichiro and Matsui. After careful calculation, WOA for Ichiro in 2004 (.317) is higher than WOA for Matsui in 2004 (.311). It may make the difference of their salaries in 2008. We remind that the Matsui was considered better than Ichiro in the comparison of OPS (.912 versus .869) and GPA (.306 versus .300).

Figure 2: The matrix plot of RPG, OPS, GPA, and WOA.

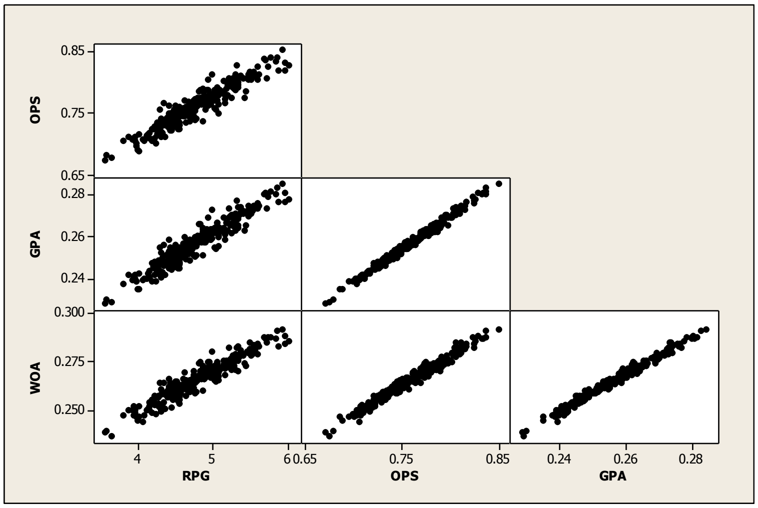


Table 7: The correlation comparison between RPG and pSLG, nP, ISO, BA, OBP, SLG, OPS, GPA, and WOA by year and by league. **Bold** represents the statistic with the highest correlation among them.

RPG vs.	TOT	AL	NL	2000	2001	2002
pSLG	0.491	0.519	0.545	0.398	0.552	0.454
nP	0.503	0.626	0.522	0.452	0.564	0.438
ISO	0.728	0.729	0.763	0.684	0.762	0.736
BA	0.786	0.766	0.767	0.797	0.843	0.830
OBP	0.882	0.898	0.878	0.942	0.926	0.836
SLG	0.895	0.886	0.903	0.870	0.879	0.915
OPS	0.946	0.947	0.952	0.929	0.934	0.935
GPA	0.951	0.955	**0.956**	0.946	0.947	0.929
WOA	**0.953**	**0.957**	**0.956**	**0.952**	**0.955**	**0.936**

RPG vs.	2003	2004	2005	2006	2007	2008
pSLG	0.661	0.582	0.539	0.359	0.277	0.479
nP	0.586	0.501	0.466	0.337	0.438	0.505
ISO	0.853	0.781	0.672	0.634	0.584	0.698
BA	0.889	0.803	0.705	0.671	0.764	0.677
OBP	0.916	0.875	0.783	0.799	0.875	0.837
SLG	0.952	0.928	0.789	0.854	0.886	0.905
OPS	0.974	**0.970**	0.879	**0.934**	0.951	0.945
GPA	0.973	**0.970**	0.894	0.933	0.958	0.943
WOA	**0.977**	**0.970**	**0.895**	**0.934**	**0.959**	**0.949**

Figure 3: The comparative box-plots of RPG and FRPG (fitted RPG on WOA) by year.

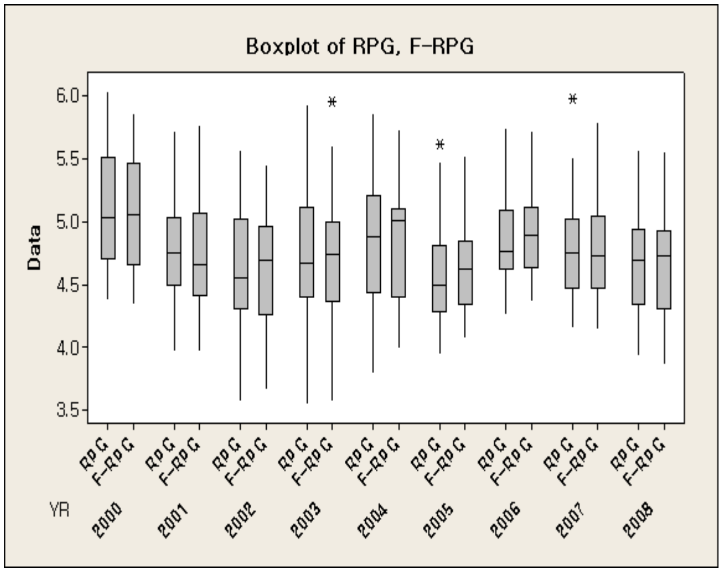


Table 8: Descriptive summary of BA, WOA, and GPA by league.

	Var	Mean	SE	Min	Q_1	Med	Q_3	Max
	BA	0.266	0.0006	0.240	0.259	0.266	0.272	0.294
TOT	WOA	0.266	0.0006	0.239	0.258	0.266	0.272	0.292
	GPA	0.257	0.0007	0.228	0.249	0.257	0.263	0.285
	Var	Mean	SE	Min	Q_1	Med	Q_3	Max
	BA	0.269	0.0009	0.240	0.263	0.269	0.277	0.290
AL	WOA	0.268	0.0010	0.239	0.259	0.269	0.275	0.292
	GPA	0.259	0.0010	0.229	0.250	0.259	0.266	0.285
	Var	Mean	SE	Min	Q_1	Med	Q_3	Max
	BA	0.263	0.0008	0.243	0.256	0.263	0.268	0.294
NL	WOA	0.264	0.0008	0.239	0.258	0.264	0.271	0.288
	GPA	0.255	0.0008	0.228	0.248	0.255	0.262	0.282

Table 9: Top 10 WOA players in 2004 along with their GPA, OPS, BA, HR, and RBIs.

Rk	Player	WOA	GPA	Rk	OPS	Rk	BA	Rk	HR	Rk	RBI	Rk
1	Pujols	0.370	0.371	1	1.114	1	0.357	2	37	4	116	9
2	Jones	0.360	0.355	2	1.044	2	0.364	1	22	59	75	78
3	Ramirez	0.345	0.344	3	1.031	3	0.332	3	37	4	121	6
4	Bradley	0.338	0.337	4	0.999	4	0.321	6	22	59	77	72
5	Berkman	0.333	0.331	5	0.986	5	0.312	11	29	29	106	17
6	Holliday	0.326	0.319	9	0.947	11	0.321	6	25	41	88	50
7	Teixeira	0.326	0.323	6	0.962	9	0.308	14	33	15	121	6
8	Rodriguez	0.324	0.320	8	0.965	7	0.302	27	35	11	103	21
9	Quentin	0.322	0.320	7	0.965	7	0.288	56	36	9	100	26
10	Youkilis	0.320	0.318	10	0.958	10	0.312	11	29	29	115	10

3. Conclusion

We have studied OPS, the sum of OBP and SLG, which has become more widely used and variations of OPS. We proposed a variation of OPS, WOA (weighted offensive average), which is a single number explaining not only a batter's hitting performance but also his non-hitting performance to generate runs for his team such as stolen bases, walks, and etc. This newly developed statistic was based on major league team statistics from the year 2000 to the year 2008. We showed WOA is the best single statistic to explain to produce runs for a team. We would like to develop salary model based on the newly developed statistic WOA. We may add the player's popularity, the indicators for salary arbitrary and free agent. And we would like to add antedependence models for the covariance.

References

Gleeman, A (2003). http://aarongleeman.com/2003_11_23_baseballblog_archive.html

James, Bill (1985). *The Bill James Historical Baseball Abstract* (1st ed.), pp. 273-4.

Kutner, M.H., C.J. Nachtsheim and J. Neter (2004). *Applied Linear Regression Models.* McGraw Hill, New York, ISBN: 978-0256086010.

Thorn, J and Palmer, P (1984), *The Hidden Game of Baseball*, pp. 69-70.

http://espn.go.com/mlb/statistics