Project 2 – R Practice Summary Report

ALY6000

Prepared By: Muhammad Umer

Presented To: Prof. John Wilder

Date: 2024/09/28

Introduction and Key Findings

Overview:

This report presents an analysis of batting statistics from the 1986 Major League Baseball season to determine potential candidates for the Most Valuable Player (MVP) award. The study aims to provide data-driven insights to support the MVP selection process, focusing on key performance metrics and their relationships.

Key Findings:

1. **Age and Performance Correlation:**

   Analysis of age-related statistics revealed interesting patterns in player performance:

	Age	Count	HR	H	R
1	20	5	3.400000	24.00000	11.80000
2	21	16	3.687500	25.18750	15.81250
3	22	35	2.514286	30.91429	15.54286
4	23	35	4.057143	39.85714	21.71429
5	24	62	4.580645	44.64516	23.16129
6	25	90	4.700000	44.74444	21.90000
7	26	84	5.833333	51.03571	25.50000
8	27	57	5.105263	57.43860	30.00000
9	28	61	4.131148	51.72131	27.09836
10	29	49	5.693878	56.87755	28.51020
11	30	45	4.244444	54.86667	25.28889
12	31	44	4.340909	42.38636	22.65909

   Figure 1: Age-related performance statistics for the 1986 MLB season.

   The data suggests that players in their mid-20s tend to have higher performance across multiple metrics, including home runs (HR), hits (H), and runs (R).

2. **Strikeout Leaders**:

Identifying players with the highest strikeout rates provides insight into batting discipline:

	Last	First	Age	G	PA	AB	R	H	X2B	X3B	HR	RBI	SB	CS	BB	SO
1	Incaviglia	Pete	22	153	606	540	82	135	21	2	30	88	3	2	55	185
2	Deer	Rob	25	134	546	466	75	108	17	3	33	86	5	2	72	179
3	Canseco	Jose	21	157	682	600	85	144	29	1	33	117	15	7	65	175
4	Presley	Jim	24	155	660	616	83	163	33	4	27	107	0	4	32	172
5	Tartabull	Danny	23	137	578	511	76	138	25	6	25	96	4	8	61	157
6	Balboni	Steve	29	138	562	512	54	117	25	1	29	88	0	0	43	146
7	Barfield	Jesse	26	158	671	589	107	170	35	2	40	108	8	8	69	146
8	Samuel	Juan	25	145	633	591	90	157	36	12	16	78	42	14	26	142
9	Murphy	Dale	30	160	692	614	89	163	29	7	29	83	7	7	75	141
10	Strawberry	Darryl	24	136	562	475	76	123	27	5	27	93	28	12	72	141

Figure 2Top 10 players with the highest strikeout rates in the 1986 MLB season.

This data highlights players who, despite high strikeout rates, may still contribute significantly in other areas such as home runs and RBIs.

3. **Home Runs vs. RBIs Correlation:**

A scatterplot analysis of home runs versus RBIs revealed a strong positive correlation between these two metrics, indicating that power hitters tend to drive in more runs.
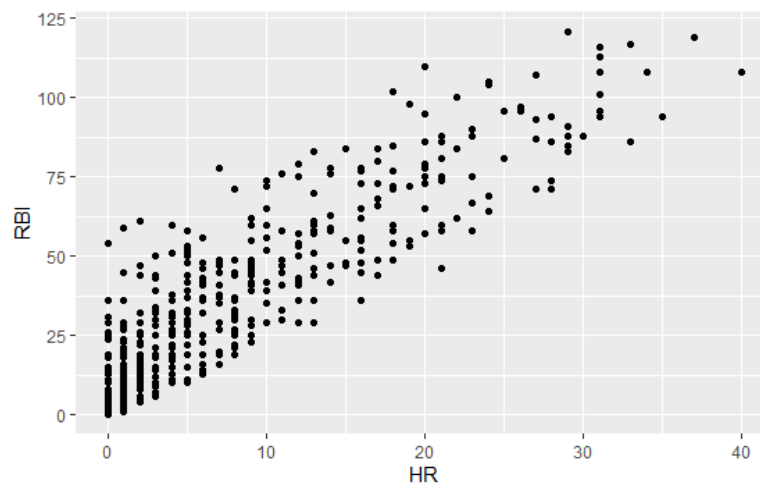

Figure 3: Scatterplot of Home Runs vs. RBIs for eligible players in the 1986 MLB season.

4. **Batting Average Distribution:**

The histogram of batting averages for eligible players provides insight into the overall hitting performance across the league:
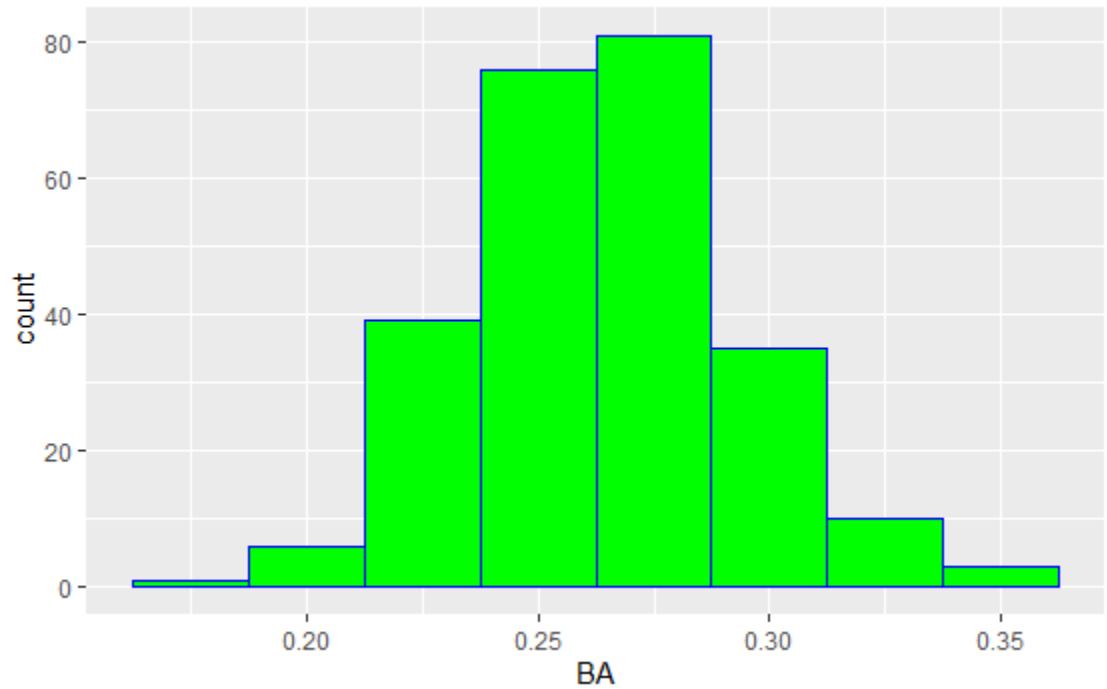

Figure 4: Distribution of batting averages for eligible players in the 1986 MLB season.

The distribution appears to be roughly normal, with a peak around the .260-.280 range, suggesting this as the typical performance level for MLB players in 1986.

5. MVP Candidates:

Based on a composite ranking of Home Runs, RBIs, and On-Base Percentage, the top MVP candidates were identified:

	First	Last	RankHR	RankRBI	RankOBP	TotalRank
1	Don	Mattingly	7	5	8	20
2	Mike	Schmidt	2	2	16	20
3	Jesse	Barfield	1	7	45	53
4	Dwight	Evans	27	17	30	74
5	Kirby	Puckett	7	18	50	75
6	Jim	Rice	52	6	18	76
7	Pete	O'Brien	36	28	17	81
8	George	Bell	7	7	74	88
9	Kevin	McReynolds	27	18	45	90
10	Kirk	Gibson	19	34	41	94
11	Gary	Gaetti	4	7	86	97
12	Von	Hayes	61	16	21	98
13	Brian	Downing	52	22	28	102
14	Darryl	Strawberry	23	26	57	106
15	Darrell	Evans	14	38	57	109
16	Kent	Hrbek	14	27	71	112
17	Eric	Davis	23	71	22	116
18	Dave	Winfield	32	12	74	118
19	Larry	Parrish	19	23	77	119
20	Eddie	Murray	74	40	6	120

Figure 5: Top 5 MVP candidates based on composite ranking of HR, RBI, and OBP.

Conclusion and Recommendations:

1. **MVP Selection**: Based on the composite ranking, Don Mattingly and Mike Schmidt emerge as the top candidates for the MVP award. Their consistent high performance across multiple offensive categories sets them apart from other players.
2. **Age Consideration**: The data suggests that players in their mid-20s tend to perform at their peak. Teams should consider this when making roster decisions and planning for player development.
3. **Balanced Evaluation**: While home runs and RBIs are important metrics, the analysis shows that on-base percentage (OBP) is also crucial. Decision-makers should consider a balanced approach when evaluating player performance.
4. **Strikeout Analysis**: High strikeout rates don't necessarily indicate poor overall performance. Players like Pete Incaviglia and Rob Deer, despite high strikeout numbers, contribute significantly in power categories. Teams should evaluate players holistically rather than focusing on a single metric.
5. **Further Analysis**: To provide a more comprehensive MVP recommendation, it is suggested that defensive metrics and pitcher performance be incorporated into future analyses. This would give a more complete picture of a player's overall value to their team.

In conclusion, while the data points to Don Mattingly and Mike Schmidt as leading MVP candidates, the final decision should consider additional factors not captured in this analysis, such as team performance and intangible leadership qualities.

Bibliography

Bluman, A. G. (2018). Elementary statistics: a step by step approach. McGraw-Hill Education.

W3 School for R Tutorials, https://www.w3schools.com/R/

Wickham, H., et al. (2024). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686