MLB HALL OF FAME PROBABILITY

HAYDEN LAMPRECHT

RESEARCH QUESTION

Do players with higher stats get into the Hall of Fame quicker?

Dependent Variable:

- Years to Hall of Fame

Independent Variables:

- WAR/Season
- All-Star Appearances

EXPLANATION OF VARIABLES

Years to Hall of Fame:

- Difference between the year a player was inducted into the hall of fame and the year the player retired

WAR/Season:

- WAR (wins above replacement) measures a player's overall value to their team
  - Taken as per season because of varying number of seasons played

All-Star Appearances:

- Each season has an All-Star game which takes the top players at each position

DATA

All data was pulled from BaseballReference.com

- Pulled top 500 WARs of all time
- Pulled complete list of Hall of Fame players
- Combined data sets
- Deleted players who were not yet retired and not in the Hall of Fame
- Deleted players whose career began before 1933 (all-star game)
- Also deleted players who do not fit current Hall of Fame qualifications
  - 10 years of experience and 5 years of retirement

DATA

- Added equation to calculate the years to Hall of Fame
- Added equation to calculate WAR/Season

Name	Inducted	Yrs	From	To	WAR/pit	WAR/pos	TOTAL WAR	WAR/season	ASG	Years to HOF
Willie Mays	1979	23	1948	1973		156.2	156.2	6.79	24	6
Henry Aaron	1982	23	1954	1976		143.1	143.1	6.22	25	6
Stan Musial	1969	22	1941	1963		128.5	128.5	5.84	24	6
Ted Williams	1966	19	1939	1960		121.8	121.8	6.41	19	6
Rickey Henderson	2009	25	1979	2003		111.1	111.1	4.44	10	6
Mickey Mantle	1974	18	1951	1968		110.2	110.2	6.12	20	6
Tom Seaver	1992	20	1967	1986	106.02	3.9	109.92	5.50	12	6
Frank Robinson	1982	21	1956	1976		107.2	107.2	5.10	14	6
Mike Schmidt	1995	18	1972	1989		106.9	106.9	5.94	12	6
Greg Maddux	2014	23	1986	2008	104.78	1.8	106.58	4.63	8	6
Randy Johnson	2015	22	1988	2009	103.52	-2.5	101.02	4.59	10	6
Joe Morgan	1990	22	1963	1984		100.6	100.6	4.57	10	6
Warren Spahn	1973	21	1942	1965	92.51	7.6	100.11	4.77	17	8
Carl Yastrzemski	1989	23	1961	1983		96.5	96.5	4.20	18	6
Cal Ripken Jr.	2007	21	1981	2001		95.9	95.9	4.57	19	6
Eddie Mathews	1978	17	1952	1968		95.9	95.9	5.64	12	10
Phil Niekro	1997	24	1964	1987	96.96	-1.1	95.86	3.99	5	10
Bert Blyleven	2011	22	1970	1992	96.12	-1.6	94.52	4.30	2	19
Adrian Beltré	2024	21	1998	2018		93.5	93.5	4.45	4	6
Al Kaline	1980	22	1953	1974		92.8	92.8	4.22	18	6
Wade Boggs	2005	18	1982	1999		91.4	91.4	5.08	12	6
Steve Carlton	1994	24	1965	1988	84.11	6.1	90.21	3.76	10	6
Gaylord Perry	1991	22	1962	1983	93.03	-3	90.03	4.09	5	8
Bob Gibson	1981	17	1959	1975	81.7	7.5	89.2	5.25	9	6
George Brett	1999	21	1973	1993		88.6	88.6	4.22	13	6

DESC STATS

WAR/season		*ASG*		*Years to HOF*	
Mean	3.649609777	Mean	9.676923077	Mean	14.37692308
Standard Error	0.096222743	Standard Error	0.385672965	Standard Error	1.090877519
Median	3.619226006	Median	9	Median	8
Mode	3.271428571	Mode	7	Mode	6
Standard Deviation	1.097108071	Standard Deviation	4.397348363	Standard Deviation	12.43791739
Sample Variance	1.20364612	Sample Variance	19.33667263	Sample Variance	154.7017889
Kurtosis	0.036146435	Kurtosis	1.68059686	Kurtosis	2.852296698
Skewness	0.363823359	Skewness	1.190361977	Skewness	1.803223018
Range	5.28601023	Range	23	Range	53
Minimum	1.505294118	Minimum	2	Minimum	6
Maximum	6.791304348	Maximum	25	Maximum	59
Sum	474.4492711	Sum	1258	Sum	1869
Count	130	Count	130	Count	130

EXPECTATIONS OF MODEL

$$\text{Years to HOF} = B_0 - B_1 \text{ WAR/Season} - B_2 \text{ All-Star Appearances} + e$$

RESULTS

Regression Statistics								
Multiple R	0.339711023							
R Square	0.115403579							
Adjusted R Square	0.101472927							
Standard Error	11.78998497							
Observations	130							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	2303.05508	1151.52754	8.28414756	0.000415259			
Residual	127	17653.47569	139.0037456					
Total	129	19956.53077						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	29.05895809	3.754801866	7.739145533	2.72015E-12	21.62888281	36.48903337	21.62888281	36.48903337
WAR/season	-3.121393723	1.022872898	-3.051594905	0.002772353	-5.145474595	-1.097312851	-5.145474595	-1.097312851
ASG	-0.340001252	0.255199729	-1.332294719	0.185149083	-0.844995461	0.164992958	-0.844995461	0.164992958

CONCLUSION

- Model is significant (F=8.28 ; p=0.0004)
- All-Star appearances is not significant
- Low R Squared (10.15%)
- Multicollinearity?
- Solutions?

Testing for Multicollinearity	
Correlation	VIF
0.37993816	1.130459017