Descriptive Statistics: Age, AvgYearlySalary

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Age	30	0	26.500	0.469	2.570	22.000	24.000	26.000	28.000
AvgYearlySalary	30	0	4349388	875342	4794444	545000	573750	2250000	6881250

Variable	Maximum
Age	32.000
AvgYearlySalary	16000000

Regression Analysis: AvgYearlySalary versus Age

The regression equation is

AvgYearlySalary = - 22147948 + 999900 Age

Model Summary

S	R-sq	R-sq (adj)
4119434	28.72%	26.18%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1.91462E+14	1.91462E+14	11.028	0.0002
Error	28	4.75153E+14	1.69697E+13		
Total	29	6.66614E+14			

Age	AvgYearlySalary
28	1275000
24	565000
26	16000000
26	7050000
25	657000
25	4166667
24	570000
24	623200
25	555000
28	6275000
24	1000000
31	2500000
29	8285714
28	575000
27	545000

24	559600
25	575000
22	548940
28	8500000
28	3125000
24	4000000
32	6750000
28	6825000
31	12500000
28	14000000
25	4333333
23	545000
30	15000000
28	2000000
25	577200

r=0.54

Sources: "MLB 2019 Payroll Tracker." *Spotrac.com*, 2018, www.spotrac.com/mlb/payroll/.
"MLB Teams." *ESPN*, ESPN Internet Ventures, 2018, www.espn.com/mlb/teams.

       The intent of this study was to determine whether there was a relationship between the age of an MLB player and his salary (if older players make more because of their experience or if younger players make more because of their skill and energy). In attempt to find the answer to this question, I collected data on the salaries of the 30 starting shortstops in the league. I chose to only look at shortstops because position can often influence salary. Each team had one starting shortstop, so these were obtained using statistics from ESPN. By "starting shortstop" it is meant that the particular player started at shortstop for the majority of that team's season. A player's salary can also depend on how many years the contract is for, so I considered average yearly salary in my data collection (data on age coincides with the year for the salary).

       The correlation coefficient for the data was 0.54, meaning there was a moderate linear correlation between the two variables. As age increased, so did the player's average yearly salary. After graphing the residuals versus fits, it can be concluded that the relationship is indeed linear and not anything else because the graph is entirely random. No pattern in the residual plot proves that the relationship is linear. Therefore,

there is a relationship between the age of an MLB player and his salary. This relationship, although it is not strong, is a positive one.