Table 2: Variable Summary Statistics

Variable	Mean	SD	Min	Max	N
ln(Salary)	15.91	0.81	13.82	17.37	587.00
WAR	3.03	2.20	-3.30	10.70	587.00
Age	29.77	3.23	22.00	40.00	587.00
Service Time	7.43	3.21	0.00	21.00	587.00
NonWhite	0.49	0.50	0.00	1.00	587.00
International	0.33	0.47	0.00	1.00	587.00
Year	2018.36	2.36	2015.00	2022.00	587.00
Hits	145.49	23.72	79.00	216.00	587.00
Home Runs	22.50	10.27	0.00	62.00	587.00
Strikeouts	116.51	31.99	38.00	219.00	587.00
Walks	56.87	22.34	13.00	145.00	587.00
AVG	0.27	0.03	0.17	0.35	587.00
SLG	0.46	0.07	0.27	0.69	587.00
OBP	0.34	0.04	0.24	0.47	587.00
OPS	0.80	0.09	0.54	1.11	587.00
xBA	0.26	0.02	0.19	0.34	587.00
xSLG	0.45	0.07	0.27	0.71	587.00
xwOBA	0.34	0.04	0.26	0.46	587.00
xOBP	0.34	0.03	0.26	0.46	587.00
xISO	0.18	0.06	0.04	0.40	587.00
Avg EV	89.13	2.11	80.50	95.90	587.00
Avg LA	12.95	4.50	-4.40	22.70	587.00
Sweet Spot %	34.23	3.83	19.90	46.40	587.00
Barrel %	7.91	4.01	0.00	26.50	587.00

if this is same for all variables, take out column & just note in a note underneath

Also mention date ranges in note underneath

When cleaning the data, I had to make important assumptions to control for various things. Since WAR is a counting stat and would be biased towards starters than bench players, I chose only to include players who qualified for awards at the end of the season. For hitters, this requires a player to have greater than or equal to 502 plate appearances. A player must have greater than or equal to 162 innings pitched for pitchers. Although this limits the study to only starting pitchers, most of the analysis and conclusions are drawn from tests ran on the hitter dataset. I then refined the dataset even more by only keeping players who were not being paid the minimum salary during those years. The minimum salary in 2015 was $507,500 and is now currently $700,000. I chose not to include these players because they had not faced a reasonable time to be discriminated against. The instrument that can cause pay discrimination is when players sign new contracts perceived to be based on performance, not race. Recent papers decided not to include players in their arbitration years, which is the first 6 years in the MLB. Arbitration still involves some level of team input into how much they believe a player is worth, which is why I did not make this distinction in my study.

"discriminat" ~~probably~~ maybe more likely to occur at initial signing. ok, though, to exclude league minimum due to censoring.

Due to the robustness of publicly available baseball data, the dataset is full and has meaningful statistics to represent a player's value they add to their teams. Thanks to the increase in data science techniques used in baseball, the perception is that teams are increasing their accuracy in player evaluation models. However, it is worth noting that when analyzing the results of these tests, they are all taken through an academic lens because it is impossible to switch races and truly learn the effect race has on salary in the MLB.

3 Methods and Results

Our main objective is to estimate the effects of race on the yearly salary for an MLB player. In the following regressions, we try to control for outside factors to accurately define that effect.

The first model is simple and only controls for WAR, age, race, and year. From table 3, you can see

Could pose some econometric issues, though