# Predicting MLB Player Salaries

Dave Oxnard

1) Is it possible to predict a player's salary, given his prior-season performance?

2) What can be inferred about the effect of different statistics on a player's level of compensation?

# The Data

9 features

3780 Rows

**Pitchers**

| Team | IP | ERA | SO | Salary |
|---|---|---|---|---|
| SFG | 200 | 2.76 | 56 | $300,000 |
| BOS | 256 | 3.55 | 98 | $550,000 |
| LAA | 435 | 1.76 | 106 | $13,000,000 |
| CHC | 189 | 10.65 | 45 | $23,000,000 |

9 features

3021 Rows

**Batters**

| Team | AVG | OBP | HR | Salary |
|---|---|---|---|---|
| BOS | .314 | .468 | 45 | $545,000 |
| LAD | .245 | .521 | 32 | $19,000,000 |
| NYY | .119 | .276 | 0 | $300,000 |
| OAK | .302 | .409 | 38 | $15,500,000 |

- Pitchers:
  - ERA
  - IP
  - SO
  - SO9
  - Wins
  - Losses
  - WAR
- Batters:
  - AVG
  - OBP
  - OPS
  - SLG
  - HR
  - RBI
  - WAR
- Other:
  - Year
  - Age
  - Salary

# The Data

9 features

3780 Rows

**Pitchers**

| Team | IP | ERA | SO | Salary |
|---|---|---|---|---|
| SFG | 200 | 2.76 | 56 | $300,000 |
| BOS | 256 | 3.55 | 98 | $550,000 |
| LAA | 435 | 1.76 | 106 | $13,000,000 |
| CHC | 189 | 10.65 | 45 | $23,000,000 |

9 features

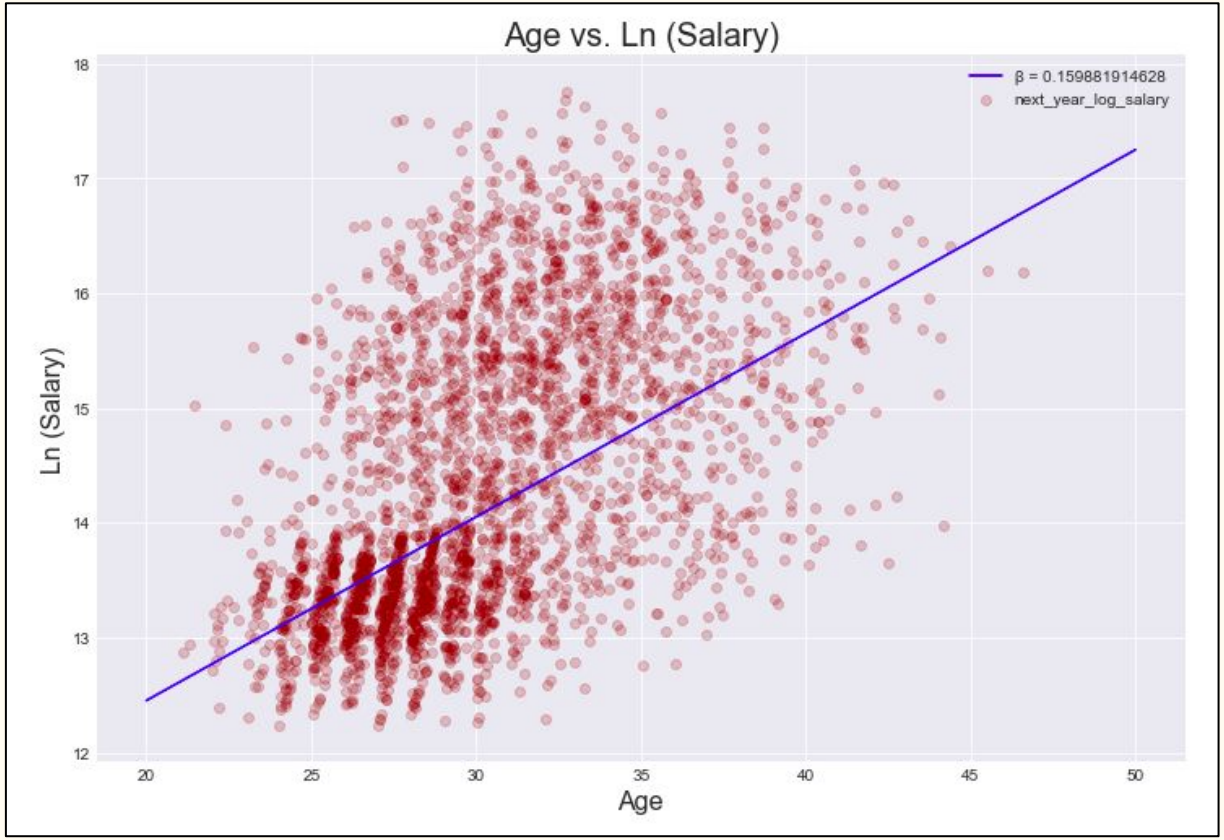
3021 Rows

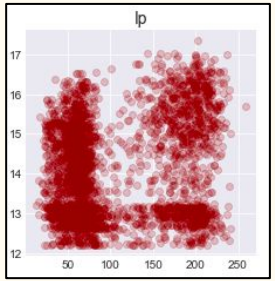
**Batters**

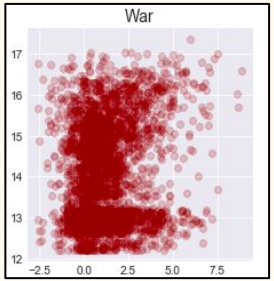
| Team | AVG | OBP | HR | Salary |
|---|---|---|---|---|
| BOS | .314 | .468 | 45 | $545,000 |
| LAD | .245 | .521 | 32 | $19,000,000 |
| NYY | .119 | .276 | 0 | $300,000 |
| OAK | .302 | .409 | 38 | $15,500,000 |

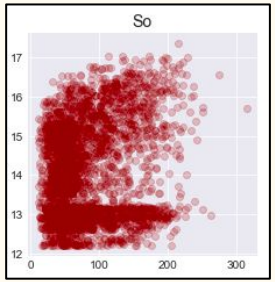
- 16 years MLB player data (2000-2016)
- Scraped with `scrapy`
- Ln (Salary) transform
- Source: baseball-reference.com

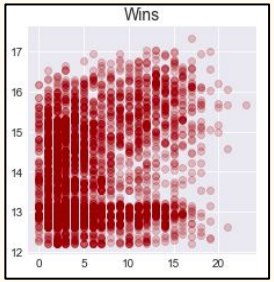
# EDA - Pitchers

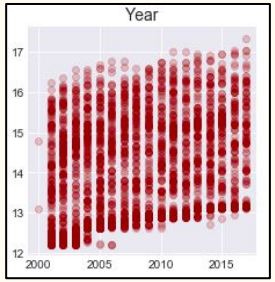


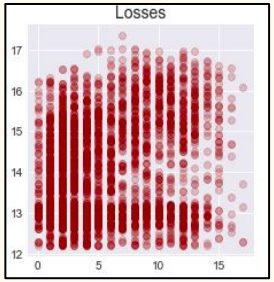



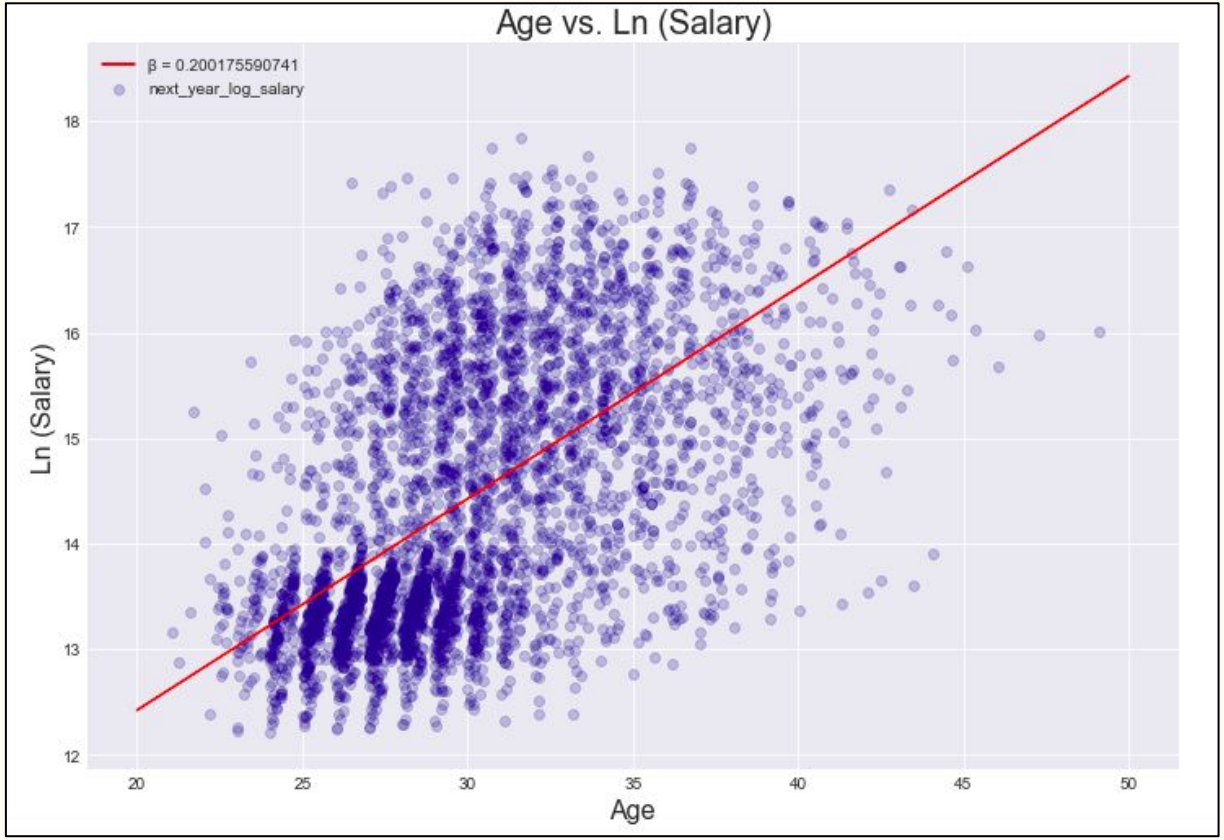



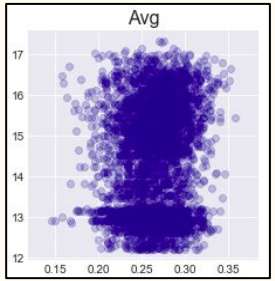



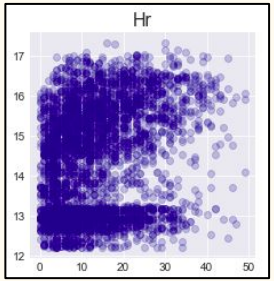



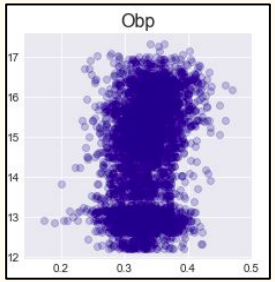



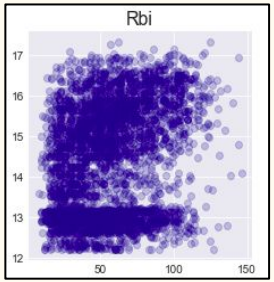
# EDA - Batters

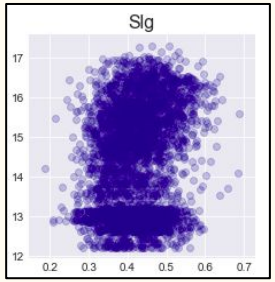

Age vs. Ln (Salary)
β = 0.200175590741
next_year_log_salary
Ln (Salary)
Age
12
13
14
15
16
17
18
20
25
30
35
40
45
50

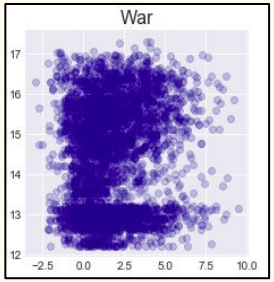

Avg
12
13
14
15
16
17
0.15
0.20
0.25
0.30
0.35


Hr
12
13
14
15
16
17
0
10
20
30
40
50


Obp
12
13
14
15
16
17
0.2
0.3
0.4
0.5


Rbi
12
13
14
15
16
17
50
100
150


Slg
12
13
14
15
16
17
0.2
0.3
0.4
0.5
0.6
0.7


War
12
13
14
15
16
17
−2.5
0.0
2.5
5.0
7.5
10.0

# “Naive” Models

## Pitchers

| | | | |
|---|---|---|---|
| **Dep. Variable:** | next_year_log_salary | **R-squared:** | 0.435 |
| **Model:** | OLS | **Adj. R-squared:** | 0.434 |
| **Method:** | Least Squares | **F-statistic:** | 290.3 |
| **Date:** | Thu, 05 Oct 2017 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 13:49:45 | **Log-Likelihood:** | -4071.4 |
| **No. Observations:** | 3021 | **AIC:** | 8161. |
| **Df Residuals:** | 3012 | **BIC:** | 8215. |
| **Df Model:** | 8 | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | -86.1722 | 7.575 | -11.377 | 0.000 | -101.024 | -71.320 |
| **age** | 0.1836 | 0.004 | 42.941 | 0.000 | 0.175 | 0.192 |
| **ip** | 0.0055 | 0.001 | 3.889 | 0.000 | 0.003 | 0.008 |
| **losses** | 0.0234 | 0.009 | 2.679 | 0.007 | 0.006 | 0.041 |
| **so** | -0.0002 | 0.001 | -0.152 | 0.879 | -0.003 | 0.003 |
| **so9** | 0.0704 | 0.016 | 4.318 | 0.000 | 0.038 | 0.102 |
| **war** | 0.0070 | 0.016 | 0.450 | 0.653 | -0.024 | 0.038 |
| **wins** | 0.0090 | 0.008 | 1.064 | 0.287 | -0.008 | 0.026 |
| **year** | 0.0465 | 0.004 | 12.318 | 0.000 | 0.039 | 0.054 |

## Batters

| | | | |
|---|---|---|---|
| **Dep. Variable:** | next_year_log_salary | **R-squared:** | 0.463 |
| **Model:** | OLS | **Adj. R-squared:** | 0.462 |
| **Method:** | Least Squares | **F-statistic:** | 361.8 |
| **Date:** | Thu, 05 Oct 2017 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 13:55:03 | **Log-Likelihood:** | -5330.3 |
| **No. Observations:** | 3780 | **AIC:** | 1.068e+04 |
| **Df Residuals:** | 3770 | **BIC:** | 1.074e+04 |
| **Df Model:** | 9 | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | -114.7636 | 6.967 | -16.472 | 0.000 | -128.424 | -101.104 |
| **age** | 0.2092 | 0.004 | 47.379 | 0.000 | 0.201 | 0.218 |
| **avg** | 1.8821 | 1.133 | 1.661 | 0.097 | -0.339 | 4.103 |
| **hr** | 0.0232 | 0.006 | 3.979 | 0.000 | 0.012 | 0.035 |
| **obp** | -47.9449 | 33.281 | -1.441 | 0.150 | -113.195 | 17.305 |
| **ops** | 52.3848 | 33.266 | 1.575 | 0.115 | -12.837 | 117.607 |
| **rbi** | 0.0121 | 0.002 | 7.752 | 0.000 | 0.009 | 0.015 |
| **slg** | -55.5395 | 33.240 | -1.671 | 0.095 | -120.709 | 9.630 |
| **war** | -0.0150 | 0.013 | -1.147 | 0.251 | -0.041 | 0.011 |
| **year** | 0.0604 | 0.003 | 17.540 | 0.000 | 0.054 | 0.067 |

# “Naive” Models

Pitchers

| | | | |
|---|---|---|---|
| Dep. Variable: | next_year_log_salary | R-squared: | 0.435 |
| Model: | OLS | Adj. R-squared: | 0.434 |
| Method: | Least Squares | F-statistic: | 290.3 |
| Date: | Thu, 05 Oct 2017 | Prob (F-statistic): | 0.00 |
| Time: | 13:49:45 | Log-Likelihood: | -4071.4 |
| No. Observations: | 3021 | AIC: | 8161. |
| Df Residuals: | 3012 | BIC: | 8215. |
| Df Model: | 8 | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | -86.1722 | 7.575 | -11.377 | 0.000 | -101.024 | -71.320 |
| age | 0.1836 | 0.004 | 42.941 | 0.000 | 0.175 | 0.192 |
| ip | 0.0055 | 0.001 | 3.889 | 0.000 | 0.003 | 0.008 |
| losses | 0.0234 | 0.009 | 2.679 | 0.007 | 0.006 | 0.041 |
| so | -0.0002 | 0.001 | -0.152 | 0.879 | -0.003 | 0.003 |
| sos | [illegible] | [illegible] | [illegible] | [illegible] | 0.038 | 0.102 |
| war | 0.0070 | 0.016 | 0.450 | 0.653 | -0.024 | 0.038 |
| wins | 0.0090 | 0.008 | 1.064 | 0.287 | -0.008 | 0.026 |
| year | 0.0465 | 0.004 | 12.318 | 0.000 | 0.039 | 0.054 |

$R^2 = 0.435$

| | | | |
|---|---|---|---|
| Dep. Variable: | next_year_log_salary | R-squared: | 0.463 |
| Model: | OLS | Adj. R-squared: | 0.462 |
| Method: | Least Squares | F-statistic: | 361.8 |
| Date: | Thu, 05 Oct 2017 | Prob (F-statistic): | 0.00 |
| Time: | 13:55:03 | Log-Likelihood: | -5330.3 |
| No. Observations: | 3780 | AIC: | 1.068e+04 |
| Df Residuals: | 3770 | BIC: | 1.074e+04 |
| Df Model: | 9 | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | -114.7636 | 6.967 | -16.472 | 0.000 | -128.424 | -101.104 |
| age | [illegible] | 0.004 | [illegible] | 0.000 | 0.201 | 0.218 |
| avg | 1.8821 | 1.133 | 1.661 | 0.097 | -0.339 | 4.103 |
| hr | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.035 |
| obp | [illegible] | [illegible] | [illegible] | [illegible] | 13.195 | 17.305 |
| ops | 52.3848 | 33.266 | 1.575 | 0.115 | -12.837 | 117.607 |
| rbi | 0.0121 | 0.002 | 7.752 | 0.000 | 0.009 | 0.015 |
| slg | -55.5395 | 33.240 | -1.671 | 0.095 | -120.709 | 9.630 |
| war | -0.0150 | 0.013 | -1.147 | 0.251 | -0.041 | 0.011 |
| year | 0.0604 | 0.003 | 17.540 | 0.000 | 0.054 | 0.067 |

**Features without statistical significance (SO, WAR, Wins)**

# “Naive” Models

Batters

| | | | |
|---|---|---|---|
| Dep. Variable: | next_year_log_salary | R-squared: | 0.463 |
| Model: | OLS | Adj. R-squared: | [illegible] |
| Method: | Least Squares | F-statistic: | 361.8 |
| Date: | Thu, 05 Oct 2017 | Prob (F-statistic): | 0.00 |
| Time: | 13:55:03 | Log-Likelihood: | -5330.3 |
| No. Observations: | 3780 | AIC: | 1.068e+04 |
| Df Residuals: | 3770 | BIC: | 1.074e+04 |
| Df Model: | 9 | | |

$R^2 = 0.463$

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | -114.7636 | 6.967 | -16.472 | 0.000 | -128.424 | -101.104 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.201 | 0.218 |
| avg | 1.8821 | 1.133 | 1.661 | 0.097 | -0.339 | 4.103 |
| hr | 0.0232 | 0.006 | 3.979 | 0.000 | 0.012 | 0.035 |
| obp | -47.9449 | 33.281 | -1.441 | 0.150 | -13.195 | 17.305 |
| ops | 52.3848 | 33.266 | 1.575 | 0.115 | -12.837 | 117.607 |
| rbi | [illegible] | [illegible] | [illegible] | [illegible] | 0.009 | 0.015 |
| slg | -55.5395 | 33.240 | -1.671 | 0.095 | -120.709 | 9.630 |
| war | -0.0150 | 0.013 | -1.147 | 0.251 | -0.041 | 0.011 |
| year | [illegible] | [illegible] | [illegible] | [illegible] | 0.054 | 0.067 |

**Features without statistical significance (AVG, OBP, OPS, WAR)**

# Final Model - Pitchers

| | | | |
|---|---|---|---|
| **Dep. Variable:** | next_year_log_salary | **R-squared:** | 0.434 |
| **Model:** | OLS | **Adj. R-squared:** | 0.433 |
| **Method:** | Least Squares | **F-statistic:** | 577.9 |
| **Date:** | Thu, 05 Oct 2017 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 14:56:02 | **Log-Likelihood:** | -4075.4 |
| **No. Observations:** | 3021 | **AIC:** | 8161. |
| **Df Residuals:** | 3016 | **BIC:** | 8191. |
| **Df Model:** | 4 | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | -86.4461 | 7.500 | -11.526 | 0.000 | -101.152 | -71.741 |
| **age** | 0.1838 | 0.004 | 43.164 | 0.000 | 0.175 | 0.192 |
| **ip** | 0.0073 | 0.000 | 25.863 | 0.000 | 0.007 | 0.008 |
| **so9** | 0.0677 | 0.009 | 7.564 | 0.000 | 0.050 | 0.085 |
| **year** | 0.0466 | 0.004 | 12.471 | 0.000 | 0.039 | 0.054 |

# Final Model - Pitchers

$R^2 = 0.434$

| | | | |
|---|---|---|---|
| Dep. Variable: | next_year_log_salary | R-squared: | 0.434 |
| Model: | OLS | Adj. R-squared: | 0.433 |
| Method: | Least Squares | F-statistic: | 577.9 |
| Date: | Thu, 05 Oct 2017 | Prob (F-statistic): | 0.00 |
| Time: | 14:56:02 | Log-Likelihood: | -4075.4 |
| No. Observations: | 3021 | AIC: | 8161. |
| Df Residuals: | 3016 | BIC: | 8191. |
| Df Model: | 4 | | |

**Unit increase in:**

**Age → 20% increase in salary**

**IP → 0.75%**

**SO9 → 7%**

**Year → 4.7%**

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | -86.4461 | 7.500 | -11.526 | 0.000 | -101.152 | -71.741 |
| age | 0.1838 | 0.004 | 43.164 | 0.000 | 0.175 | 0.192 |
| ip | 0.0073 | 0.000 | 25.863 | 0.000 | 0.007 | 0.008 |
| so9 | 0.0677 | 0.009 | 7.564 | 0.000 | 0.050 | 0.085 |
| year | 0.0466 | 0.004 | 12.471 | 0.000 | 0.039 | 0.054 |

**Features carry statistical significance**

# Final Model - Batters

| | | | |
|---|---|---|---|
| **Dep. Variable:** | next_year_log_salary | **R-squared:** | 0.459 |
| **Model:** | OLS | **Adj. R-squared:** | 0.458 |
| **Method:** | Least Squares | **F-statistic:** | 800.3 |
| **Date:** | Thu, 05 Oct 2017 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 14:35:55 | **Log-Likelihood:** | -5346.3 |
| **No. Observations:** | 3780 | **AIC:** | 1.070e+04 |
| **Df Residuals:** | 3775 | **BIC:** | 1.073e+04 |
| **Df Model:** | 4 | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | -117.8322 | 6.856 | -17.187 | 0.000 | -131.274 | -104.391 |
| **age** | 0.2118 | 0.004 | 49.561 | 0.000 | 0.203 | 0.220 |
| **rbi** | 0.0148 | 0.001 | 21.077 | 0.000 | 0.013 | 0.016 |
| **obp** | 2.5671 | 0.530 | 4.842 | 0.000 | 1.528 | 3.607 |
| **year** | 0.0618 | 0.003 | 18.212 | 0.000 | 0.055 | 0.068 |

# Final Model - Batters

$R^2 = 0.459$

| | | | |
|---|---|---|---|
| Dep. Variable: | next_year_log_salary | R-squared: | 0.459 |
| Model: | OLS | Adj. R-squared: | 0.458 |
| Method: | Least Squares | F-statistic: | 800.3 |
| Date: | Thu, 05 Oct 2017 | Prob (F-statistic): | 0.00 |
| Time: | 14:35:55 | Log-Likelihood: | -5346.3 |
| No. Observations: | 3780 | AIC: | 1.070e+04 |
| Df Residuals: | 3775 | BIC: | 1.073e+04 |
| Df Model: | 4 | | |

**Unit increase in:**

**Age → 23% increase in salary**

**RBI → 1.5%**

**OBP → 1.2%**

**Year → 6.3%**

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | -117.8322 | 6.856 | -17.187 | 0.000 | -131.274 | -104.391 |
| age | 0.2118 | 0.004 | 49.561 | 0.000 | 0.203 | 0.220 |
| rbi | 0.0148 | 0.001 | 21.077 | 0.000 | 0.013 | 0.016 |
| obp | 2.5671 | 0.530 | 4.842 | 0.000 | 1.528 | 3.607 |
| year | 0.0618 | 0.003 | 18.212 | 0.000 | 0.055 | 0.068 |

**Features carry statistical significance**

# Conclusions and Conjectures

- Statistics do not explain majority of salary variance
- Year, Age contribute significantly to salary
- Statistics are not perfect proxy for how players are valued
- Evidence of wage stickiness?
- To what extent is a player's salary a function of available talent?

# Thank You

# Appendix

# “Naive” Models

# “Naive” Models

## Pitchers

| | | | |
|---|---|---|---|
| **Dep. Variable:** | next_year_log_salary | **R-squared:** | 0.435 |
| **Model:** | OLS | **Adj. R-squared:** | 0.434 |
| **Method:** | Least Squares | **F-statistic:** | 290.3 |
| **Date:** | Thu, 05 Oct 2017 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 13:49:45 | **Log-Likelihood:** | -4071.4 |
| **No. Observations:** | 3021 | **AIC:** | 8161. |
| **Df Residuals:** | 3012 | **BIC:** | 8215. |
| **Df Model:** | 8 | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | -86.1722 | 7.575 | -11.377 | 0.000 | -101.024 | -71.320 |
| **age** | 0.1836 | 0.004 | 42.941 | 0.000 | 0.175 | 0.192 |
| **ip** | 0.0055 | 0.001 | 3.889 | 0.000 | 0.003 | 0.008 |
| **losses** | 0.0234 | 0.009 | 2.679 | 0.007 | 0.006 | 0.041 |
| **so** | -0.0002 | 0.001 | -0.152 | 0.879 | -0.003 | 0.003 |
| **so9** | 0.0704 | 0.016 | 4.318 | 0.000 | 0.038 | 0.102 |
| **war** | 0.0070 | 0.016 | 0.450 | 0.653 | -0.024 | 0.038 |
| **wins** | 0.0090 | 0.008 | 1.064 | 0.287 | -0.008 | 0.026 |
| **year** | 0.0465 | 0.004 | 12.318 | 0.000 | 0.039 | 0.054 |

## Batters

| | | | |
|---|---|---|---|
| **Dep. Variable:** | next_year_log_salary | **R-squared:** | 0.463 |
| **Model:** | OLS | **Adj. R-squared:** | 0.462 |
| **Method:** | Least Squares | **F-statistic:** | 361.8 |
| **Date:** | Thu, 05 Oct 2017 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 13:55:03 | **Log-Likelihood:** | -5330.3 |
| **No. Observations:** | 3780 | **AIC:** | 1.068e+04 |
| **Df Residuals:** | 3770 | **BIC:** | 1.074e+04 |
| **Df Model:** | 9 | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | -114.7636 | 6.967 | -16.472 | 0.000 | -128.424 | -101.104 |
| **age** | 0.2092 | 0.004 | 47.379 | 0.000 | 0.201 | 0.218 |
| **avg** | 1.8821 | 1.133 | 1.661 | 0.097 | -0.339 | 4.103 |
| **hr** | 0.0232 | 0.006 | 3.979 | 0.000 | 0.012 | 0.035 |
| **obp** | -47.9449 | 33.281 | -1.441 | 0.150 | -113.195 | 17.305 |
| **ops** | 52.3848 | 33.266 | 1.575 | 0.115 | -12.837 | 117.607 |
| **rbi** | 0.0121 | 0.002 | 7.752 | 0.000 | 0.009 | 0.015 |
| **slg** | -55.5395 | 33.240 | -1.671 | 0.095 | -120.709 | 9.630 |
| **war** | -0.0150 | 0.013 | -1.147 | 0.251 | -0.041 | 0.011 |
| **year** | 0.0604 | 0.003 | 17.540 | 0.000 | 0.054 | 0.067 |

# Final Models

# Final Model - Pitchers

| | | | |
|---|---|---|---|
| **Dep. Variable:** | next_year_log_salary | **R-squared:** | 0.434 |
| **Model:** | OLS | **Adj. R-squared:** | 0.433 |
| **Method:** | Least Squares | **F-statistic:** | 577.9 |
| **Date:** | Thu, 05 Oct 2017 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 14:56:02 | **Log-Likelihood:** | -4075.4 |
| **No. Observations:** | 3021 | **AIC:** | 8161. |
| **Df Residuals:** | 3016 | **BIC:** | 8191. |
| **Df Model:** | 4 | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | -86.4461 | 7.500 | -11.526 | 0.000 | -101.152 | -71.741 |
| **age** | 0.1838 | 0.004 | 43.164 | 0.000 | 0.175 | 0.192 |
| **ip** | 0.0073 | 0.000 | 25.863 | 0.000 | 0.007 | 0.008 |
| **so9** | 0.0677 | 0.009 | 7.564 | 0.000 | 0.050 | 0.085 |
| **year** | 0.0466 | 0.004 | 12.471 | 0.000 | 0.039 | 0.054 |

# Final Model - Batters

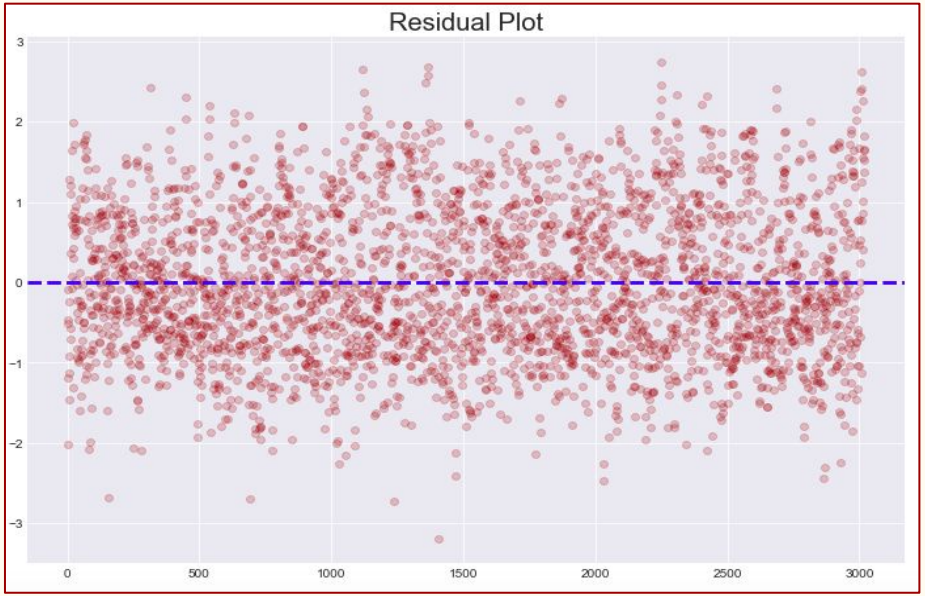
| | | | |
|---|---|---|---|
| **Dep. Variable:** | next_year_log_salary | **R-squared:** | 0.459 |
| **Model:** | OLS | **Adj. R-squared:** | 0.458 |
| **Method:** | Least Squares | **F-statistic:** | 800.3 |
| **Date:** | Thu, 05 Oct 2017 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 14:35:55 | **Log-Likelihood:** | -5346.3 |
| **No. Observations:** | 3780 | **AIC:** | 1.070e+04 |
| **Df Residuals:** | 3775 | **BIC:** | 1.073e+04 |
| **Df Model:** | 4 | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Intercept** | -117.8322 | 6.856 | -17.187 | 0.000 | -131.274 | -104.391 |
| **age** | 0.2118 | 0.004 | 49.561 | 0.000 | 0.203 | 0.220 |
| **rbi** | 0.0148 | 0.001 | 21.077 | 0.000 | 0.013 | 0.016 |
| **obp** | 2.5671 | 0.530 | 4.842 | 0.000 | 1.528 | 3.607 |
| **year** | 0.0618 | 0.003 | 18.212 | 0.000 | 0.055 | 0.068 |

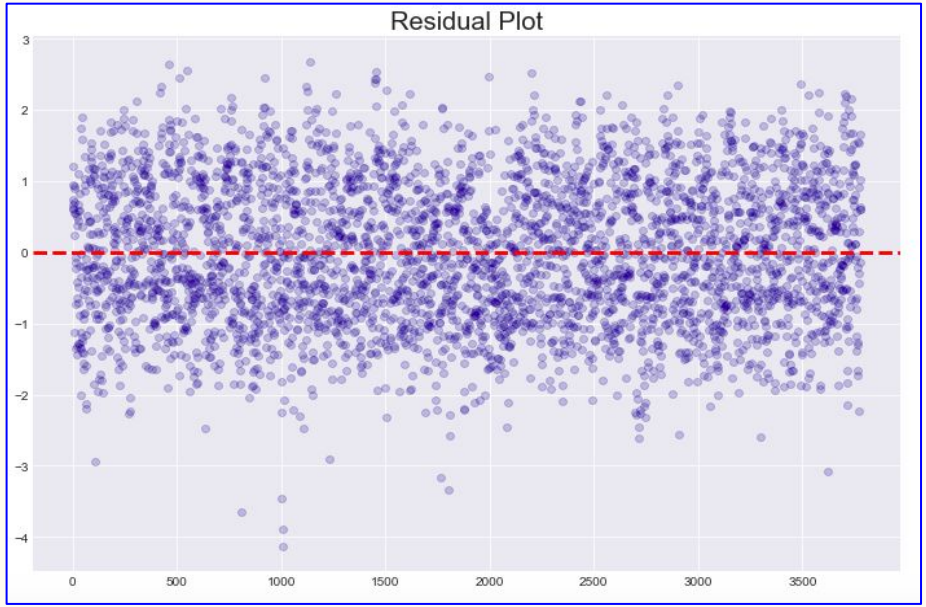
# Residual Plots

# Residual Plots

Pitchers

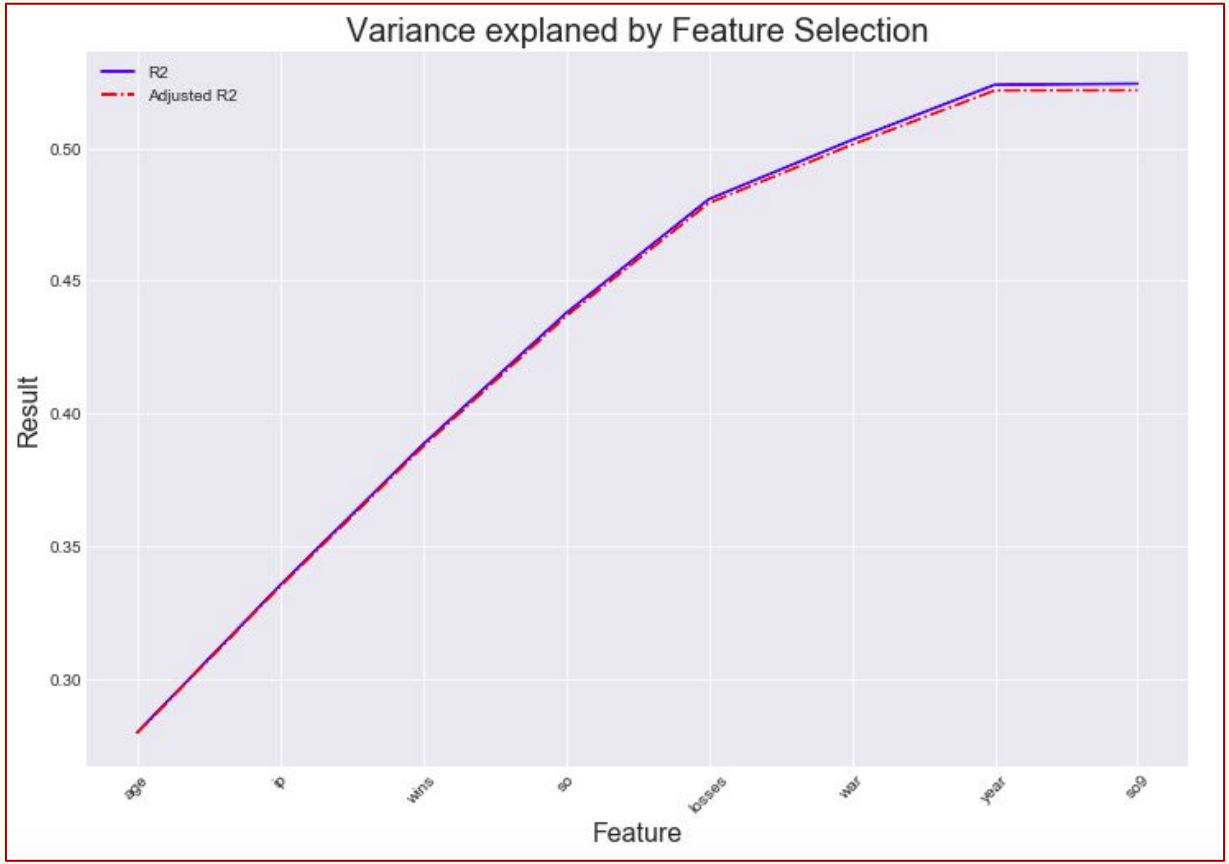



Batters

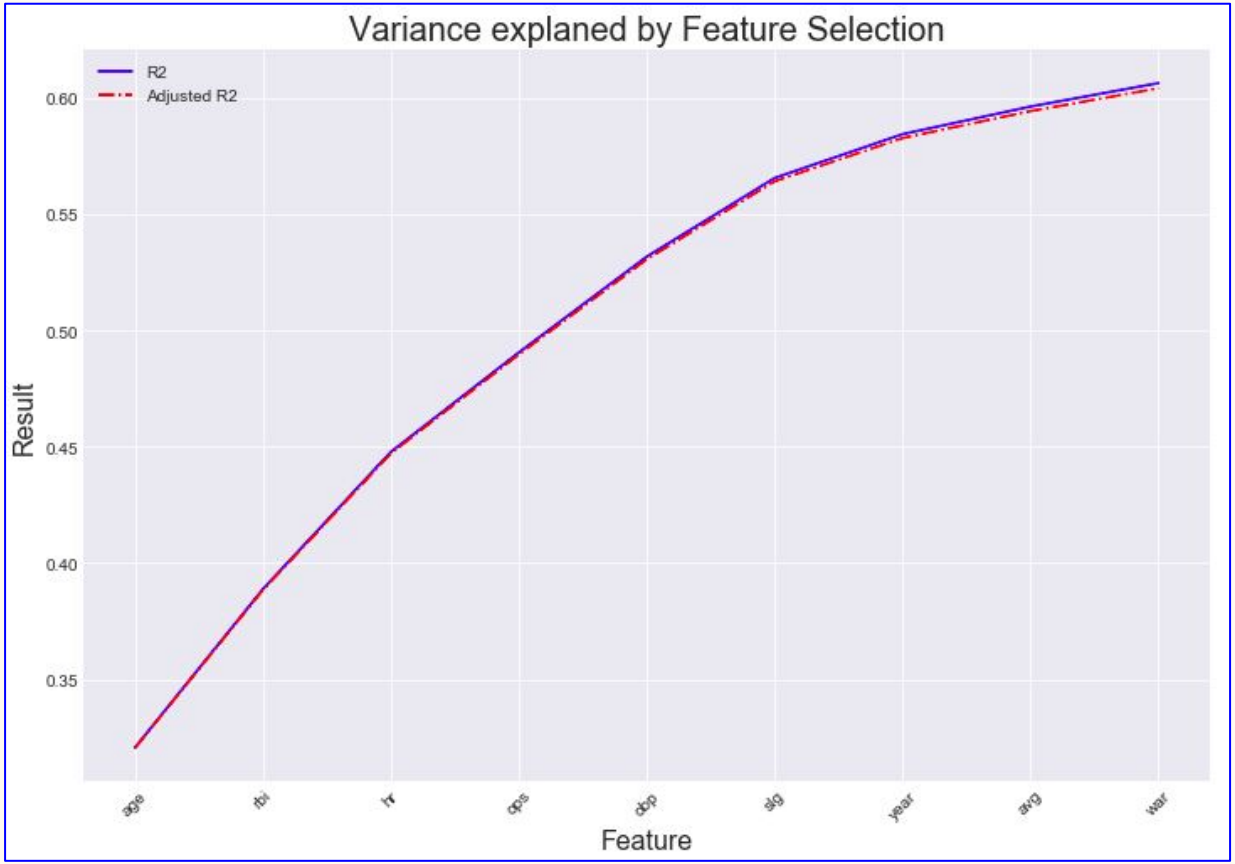


# Feature Selection

# Feature Selection - Pitchers

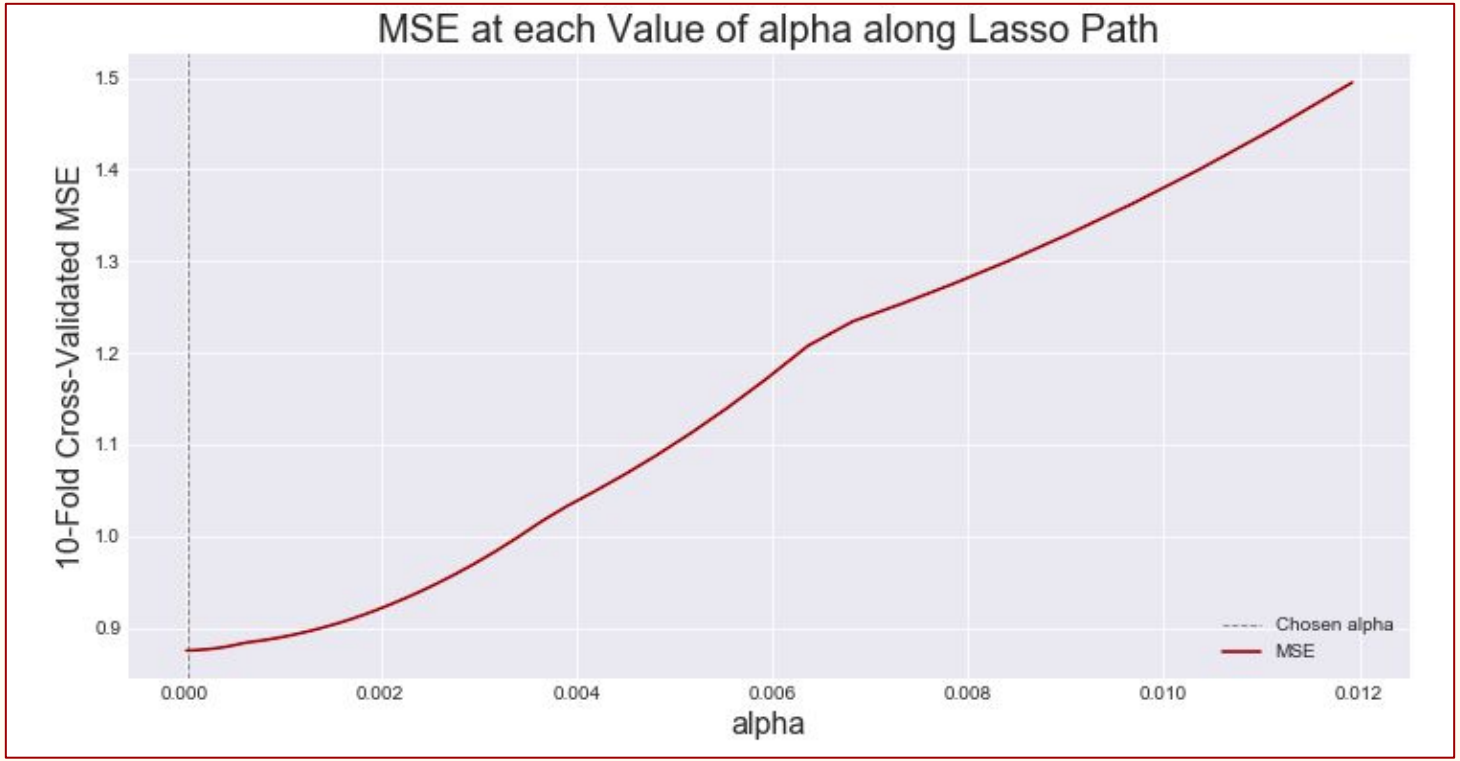


# Feature Selection - Batters

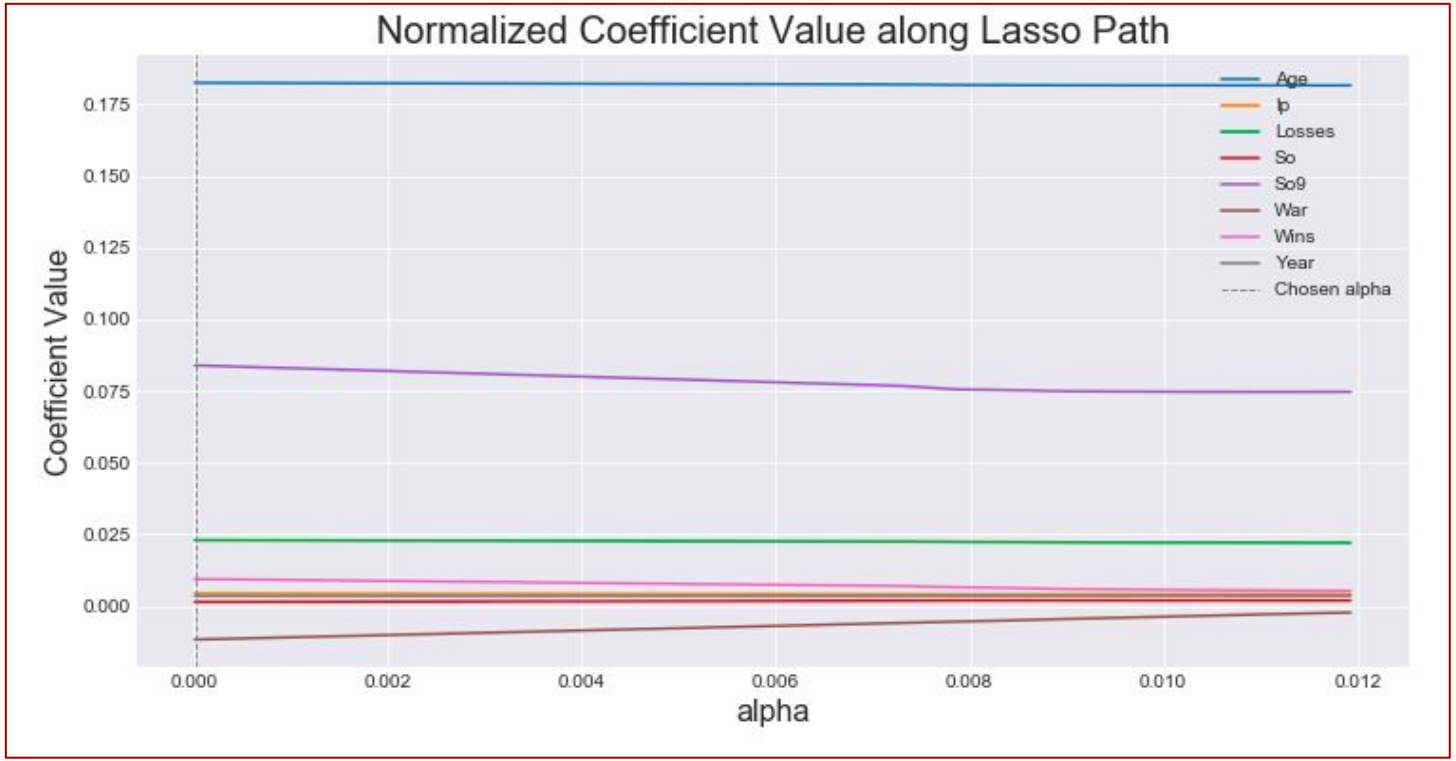


# Cross-Validation with Lasso

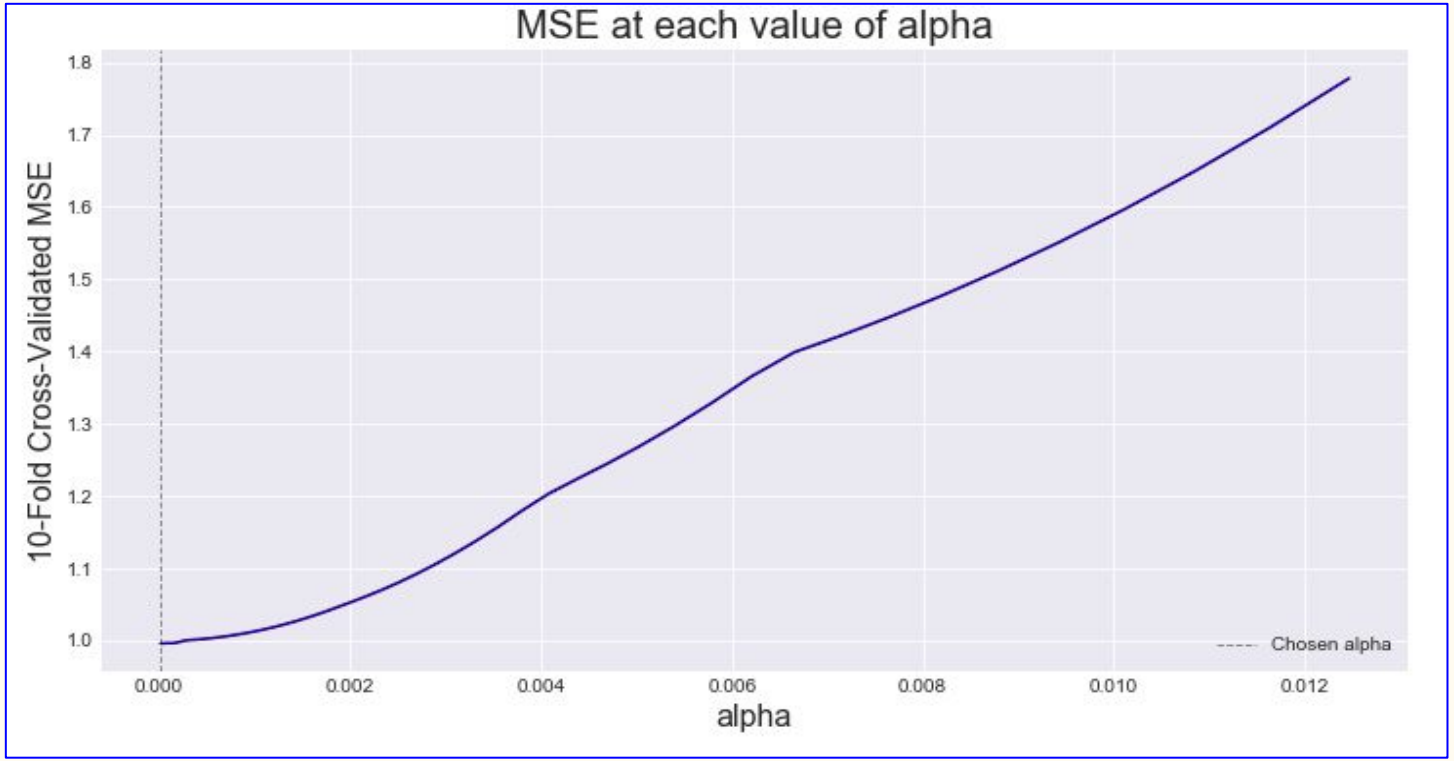
# Cross-Validation - Pitchers



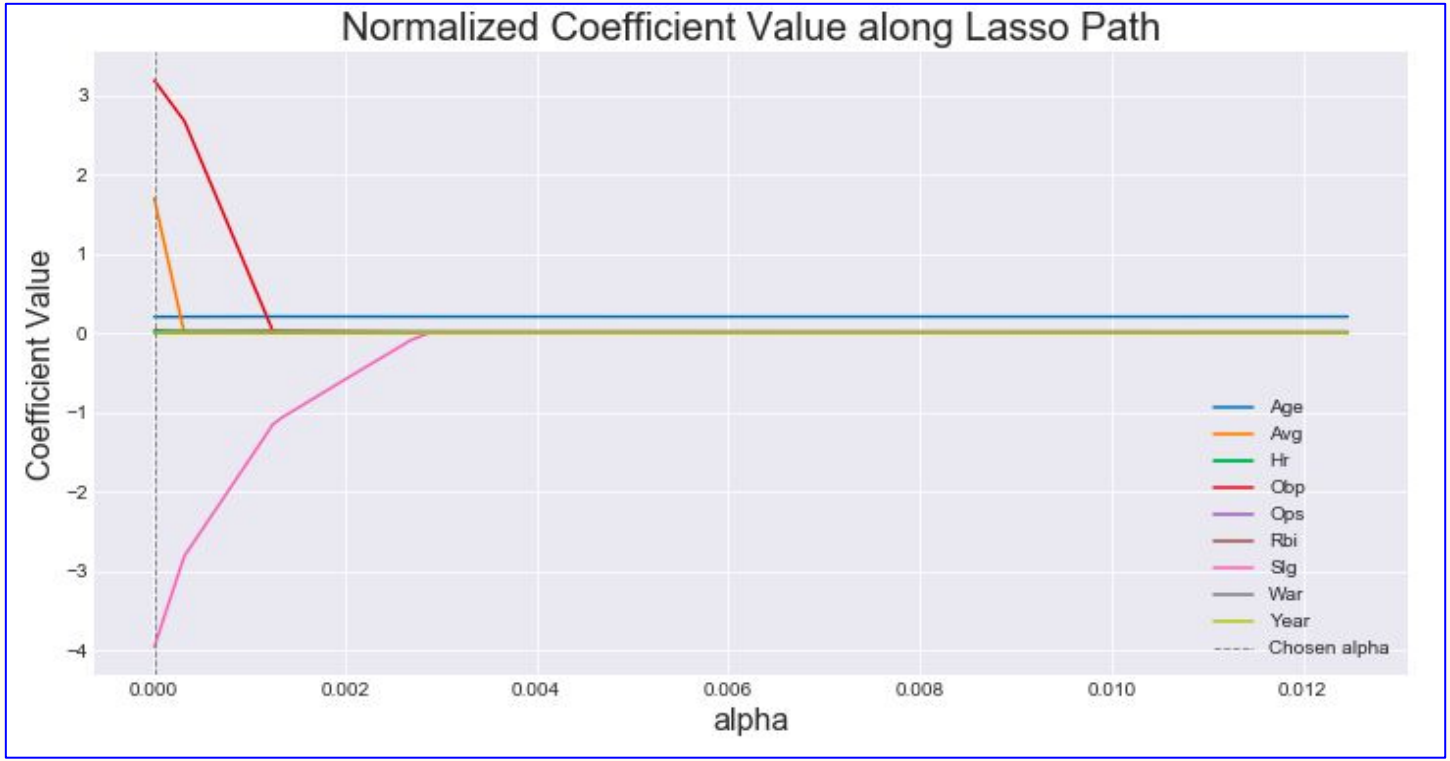
# Cross-Validation - Pitchers



# Cross-Validation - Batters



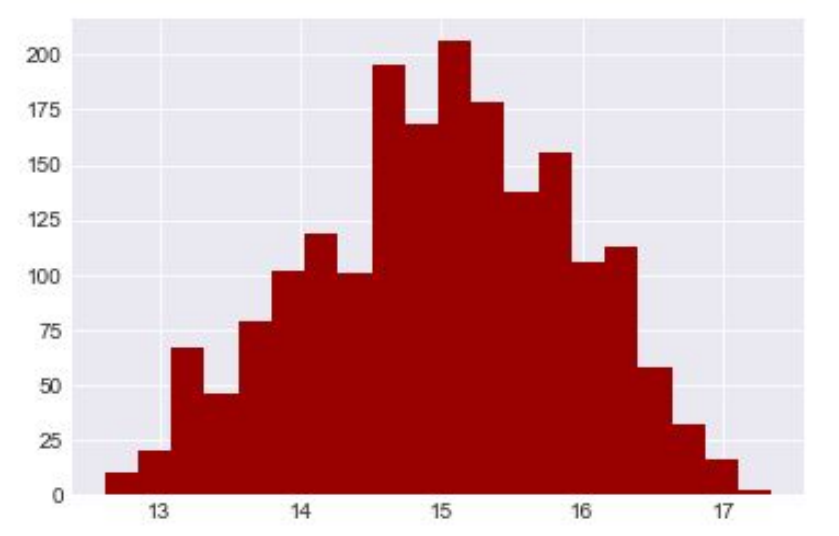
# Cross-Validation - Pitchers

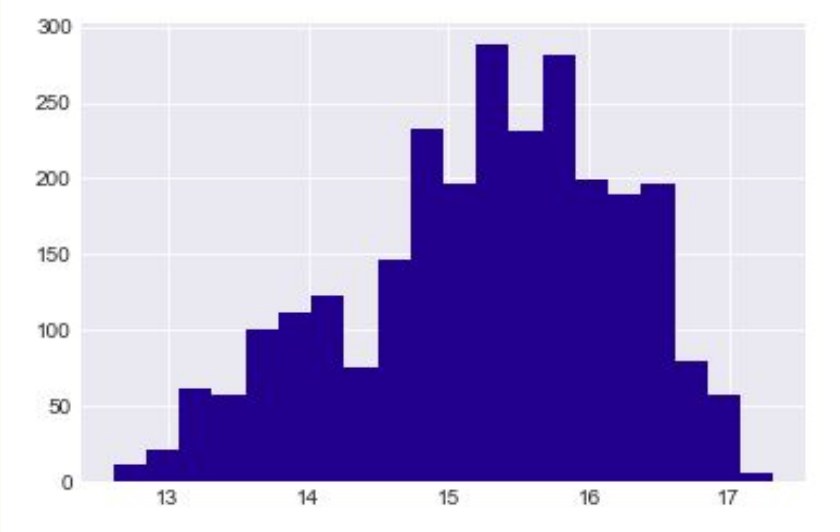


# Things I tried unsuccessfully

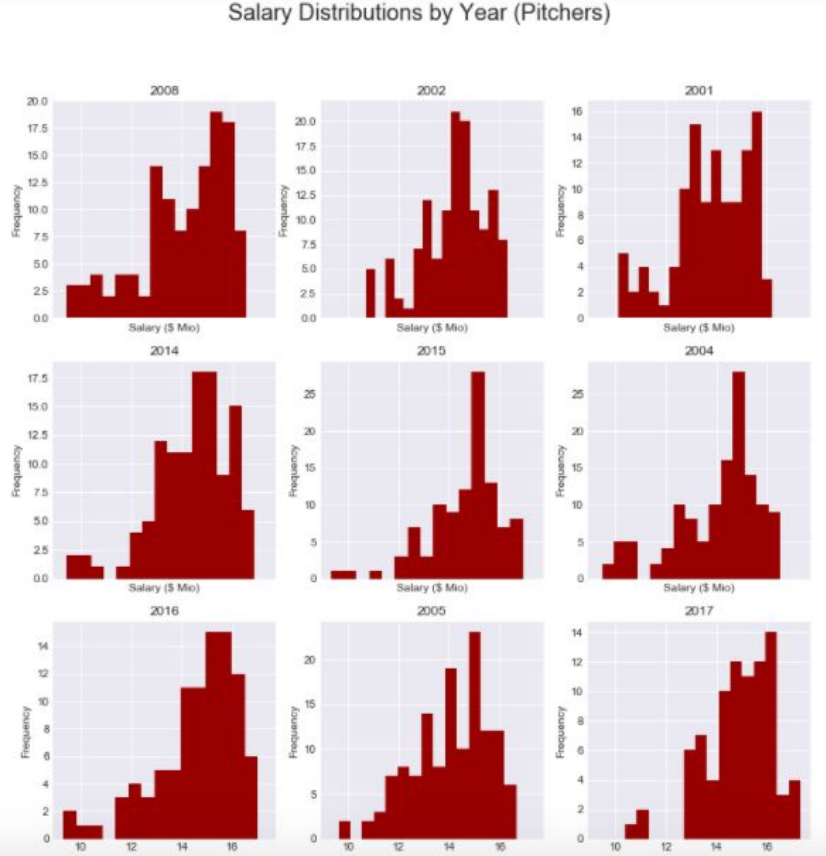
# Examining a Subset of Players

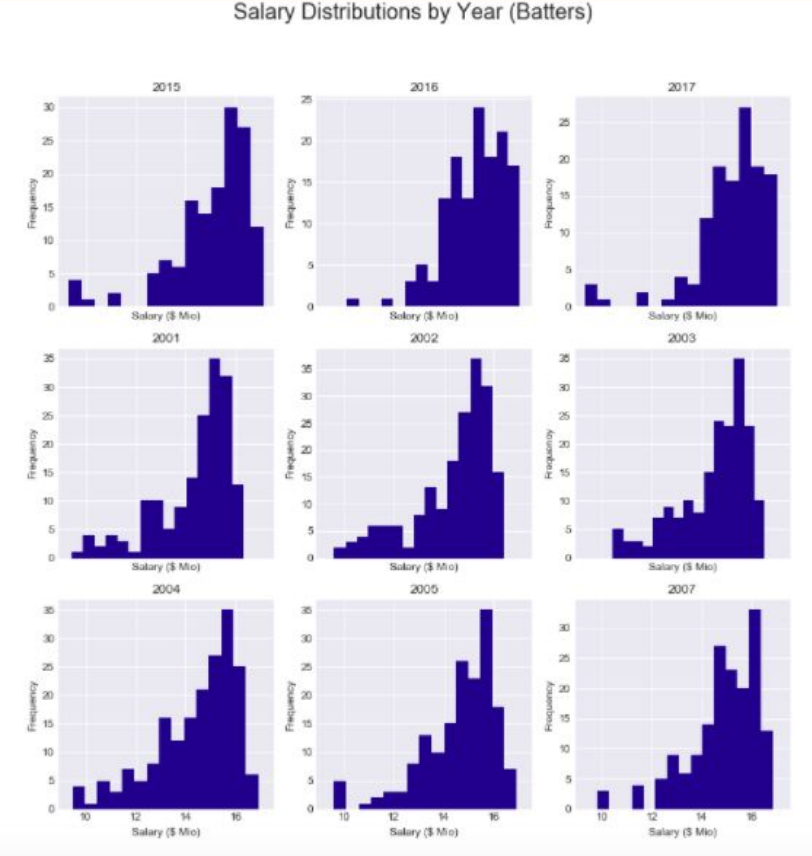
Histograms of Ln (Salary) data for players earning more than the league minimum



# Examining a Subset of Players



# Examining a Subset of Years

Pitchers

| | | | |
|---|---|---|---|
| **Dep. Variable:** | salary_over_minimum | **R-squared:** | 0.331 |
| **Model:** | OLS | **Adj. R-squared:** | 0.315 |
| **Method:** | Least Squares | **F-statistic:** | 19.76 |
| **Date:** | Wed, 04 Oct 2017 | **Prob (F-statistic):** | 3.53e-24 |
| **Time:** | 18:10:29 | **Log-Likelihood:** | -5447.6 |
| **No. Observations:** | 328 | **AIC:** | 1.091e+04 |
| **Df Residuals:** | 319 | **BIC:** | 1.095e+04 |
| **Df Model:** | 8 | | |

Batters

| | | | |
|---|---|---|---|
| **Dep. Variable:** | salary_over_minimum | **R-squared:** | 0.319 |
| **Model:** | OLS | **Adj. R-squared:** | 0.304 |
| **Method:** | Least Squares | **F-statistic:** | 22.19 |
| **Date:** | Wed, 04 Oct 2017 | **Prob (F-statistic):** | 6.32e-31 |
| **Time:** | 18:14:59 | **Log-Likelihood:** | -7335.3 |
| **No. Observations:** | 437 | **AIC:** | 1.469e+04 |
| **Df Residuals:** | 427 | **BIC:** | 1.473e+04 |
| **Df Model:** | 9 | | |

# Ideas for Future Study

# Possible Extensions

- Restrict analysis to players in salary negotiation year
- Examine effects on longer time scale (multi-season moving average)
- Control vis-a-vis team-specific factors
- Combine multiple data sources (non-stats factors)