moneyball

Matthew Wankiewicz

26/11/2020

Abstract

Abstract goes here

Keywords: Moneyball, Baseball, Sabermetrics, Logistic Regression

Introduction

This data analysis was conducted using R (R Core Team (2019)), and in particular the packages Tidyverse (Wickham et al. (2019)) and Lahman (Friendly et al. (2020)) and was compiled using R markdown (Xie, Allaire, and Grolemund (2018)).

Data

Model

```
##
## Call:
  glm(formula = playoffs ~ HR + ERA + `Est. Payroll`, family = binomial(),
       data = all_years)
##
## Deviance Residuals:
##
      Min
                1Q
                     Median
                                   3Q
                                           Max
## -2.1115 -0.5674 -0.3114 -0.0942
                                        2.8845
##
## Coefficients:
##
                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)
                   4.955e+00 5.276e-01
                                          9.392 < 2e-16 ***
## HR
                   2.415e-02
                             1.925e-03 12.549
                                                 < 2e-16 ***
                                                < 2e-16 ***
## ERA
                  -2.645e+00
                             1.630e-01 -16.229
## `Est. Payroll`
                  8.371e-09
                             1.407e-09
                                         5.949 2.69e-09 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
##
  (Dispersion parameter for binomial family taken to be 1)
##
      Null deviance: 2037.6 on 2095
                                       degrees of freedom
## Residual deviance: 1447.4 on 2092
                                      degrees of freedom
## AIC: 1455.4
## Number of Fisher Scoring iterations: 6
```

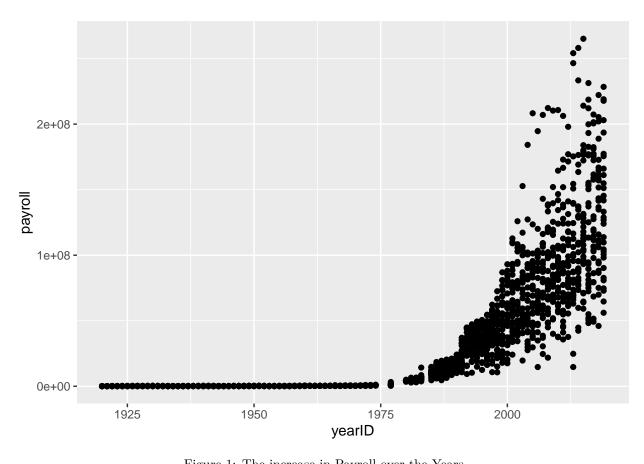


Figure 1: The increase in Payroll over the Years

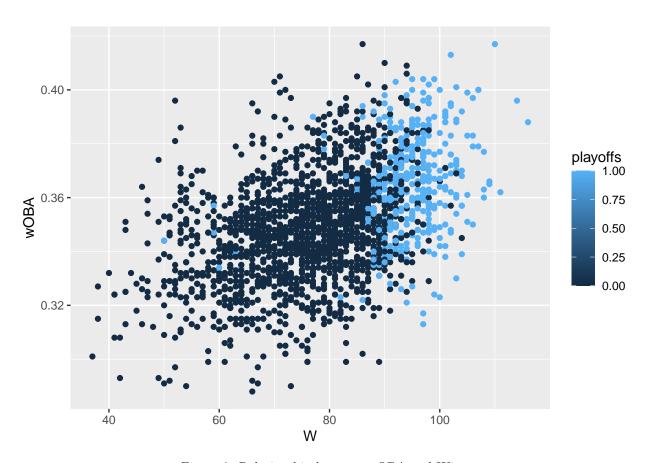


Figure 2: Relationship between wOBA and Wins

Discussion

Weaknesses and Next Steps

References

Friendly, Michael, Chris Dalzell, Martin Monkman, and Dennis Murphy. 2020. Lahman: Sean 'Lahman' Baseball Database. https://CRAN.R-project.org/package=Lahman.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui, J.J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown.