Probability and Statistics

Will Pauls

February 19, 2019

## Probability and Statistics

### Wins and Payroll Rank

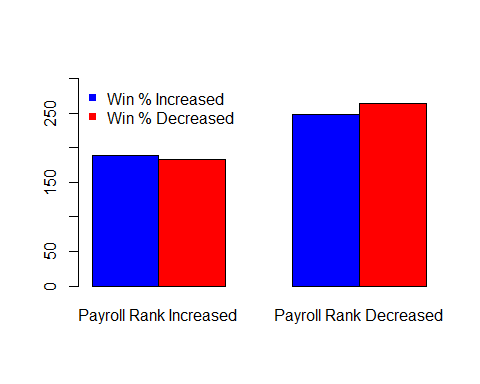
We are looking to see if increasing where you rank in terms of team payroll is correlated with more winning. (This does not necessarily mean the amount a team paid in Payroll went up.)

# summary of teams whose winning increased after payroll rank increased  
summary(baseball$winincreased[baseball$payrankincreased=="Yes"])

## No Yes NA's   
## 183 189 34

# summary of teams whose winning increased after payroll rank decreased  
summary(baseball$winincreased[baseball$payrankincreased=="No"])

## No Yes NA's   
## 264 248 34



#### Hypothesis

: = - The proportion of teams whose wins increased after increasing their Payroll Rank equals the proportion of teams whose winning increased after their payroll rank decreased.  
: - The proportion of teams whose wins increased after increasing their Payroll Rank is greater than The proportion of teams whose winning increased after their payroll rank decreased.

payranktest <- prop.test(x = c(189, 248), n = c((189+183), (248+264)), alternative = "greater")  
# Printing the results  
payranktest

##   
## 2-sample test for equality of proportions with continuity  
## correction  
##   
## data: c(189, 248) out of c((189 + 183), (248 + 264))  
## X-squared = 0.39359, df = 1, p-value = 0.2652  
## alternative hypothesis: greater  
## 95 percent confidence interval:  
## -0.03464492 1.00000000  
## sample estimates:  
## prop 1 prop 2   
## 0.5080645 0.4843750

Since the p-value is .2652 we do not have enough evidence to reject the null hypothesis that the two proportions are equal to each other.

### Wins and Payroll Amount

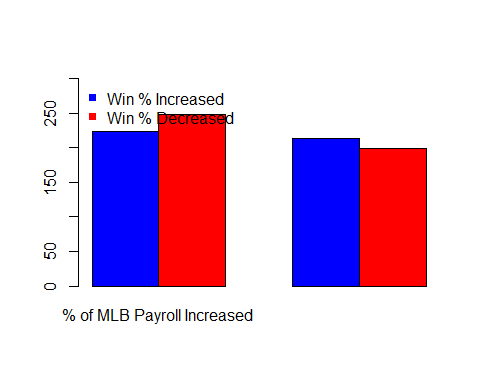
We are looking to see if increasing the amount of money you spend in Payroll in relation to other teams is correlated with winning

# summary of teams whose winning increased after % of mlb pay increased  
summary(baseball$winincreased[baseball$percentofmlbincreased=="Yes"])

## No Yes NA's   
## 248 224 34

# summary of teams whose winning increased after % of mlb pay decreased  
summary(baseball$winincreased[baseball$percentofmlbincreased=="No"])

## No Yes NA's   
## 199 213 34



#### Hypothesis

Since a lower proportion of teams have increased their winning after increasing their payroll than when they decrease in relation to other teams we will test both sides for this hypothesis.

: = - The proportion of teams whose wins increased after increasing their Payroll amount in relation to other MLB teams equals the proportion of teams whose winning increased after their payroll amount decreased in relation to other teams.  
: - The proportion of teams whose wins increased after increasing their Payroll amount in relation to other MLB teams does not equal the proportion of teams whose winning increased after their payroll amount decreased in relation to other teams.

payrolltest <- prop.test(x = c(224, 213), n = c((224+248), (213+199)))  
# Printing the results  
payrolltest

##   
## 2-sample test for equality of proportions with continuity  
## correction  
##   
## data: c(224, 213) out of c((224 + 248), (213 + 199))  
## X-squared = 1.418, df = 1, p-value = 0.2337  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## -0.11069982 0.02587178  
## sample estimates:  
## prop 1 prop 2   
## 0.4745763 0.5169903

Since the p-value is .2337 we do not have enough evidence to reject the null hypothesis that the two proportions are equal to each other.

### Conclusion

Looking at the proportions of teams whose winning increased after they increased their rank in payroll and after they increased their % of MLB payroll, it is apparent that the amount of money a team puts into their payroll isn't necessarily as important as where they rank in team payroll regardless of how much they spend. For this reason, our linear regression model we will use in the Machine Learning section will use Payroll Rank as the Independent variable.