MLB Payroll and Team Performance: An Economic Data Analysis

https://colab.research.google.com/drive/1D5JpOcTtcuglwdFnEC3CCayRf4Spqlfo?usp=sharing

This project asks a straightforward economic question: **do MLB teams that spend more on payroll win more games?** Payroll is a team's main investment in player talent, so it makes sense to check whether higher spending actually shows up in better performance. MLB is a useful setting for this because payroll varies a lot across teams, giving us enough variation to measure a relationship. The goal is not to explain everything that affects wins, but to estimate how strongly payroll is connected to success in a given season.

We took our data from a Kaggle dataset, which had data from 2012-2024 MLB payrolls and average attendance. We dropped all of the values in the 2020 year, because all teams had zero attendance due to COVID. We also dropped a couple other values that did not have attendance values, like the 2024 Oakland Athletics. We also have a couple other datasets that we used to create bins for the attendance figures, and another to further limit the data to teams who averaged 10-20 thousand fans and those that averaged over 40,000 fans. We used scatterplots for all of our data points, except for playoff success which uses a side by side bar graph (which we changed from our presentation). The graphs, models, and residuals are below.

To study this relationship, we estimated a **multiple linear regression model** using total wins (W_Record) as the outcome. The model includes two predictors: payroll share and average attendance. Payroll share measures how much each team contributes to total league payroll, while attendance acts as a control variable that captures factors like market size, revenue strength, and overall organizational support. These two variables together allow us to separate the effect of spending from the effect of having a stronger or larger fanbase.

The model is written as:

$$Wins = \beta 0 + \beta 1 \cdot PayrollPct + \beta 2 \cdot AverageAttendancei + \varepsilon$$

In this equation, $\beta 0$ is the baseline number of wins when both predictors are zero. $\beta 1$ measures how wins change when payroll share increases by one percentage point, holding

attendance constant. $\beta 2$ measures how wins change with increases in attendance, holding payroll share constant. The error term ε captures everything else that affects wins but is not included in the model, such as injuries, roster decisions, coaching, or luck. This setup allows us to isolate the association between spending and performance rather than mixing it with other factors.

The regression produced an intercept of **62.9053**, meaning the model predicts around 63 wins at a starting point before payroll and attendance effects are added. The coefficient on payroll share is **2.0586**, with a p-value of 0.001. This means that **for every one–percentage-point increase in payroll share, a team is predicted to win about two additional games**, holding attendance constant. The coefficient on average attendance is **0.4397**, also statistically significant, showing that teams with higher attendance tend to win more games even after controlling for spending. Both predictors move wins in the expected direction: more spending and stronger attendance are each associated with better performance.

Although we did not explore it statistically, we also investigated how teams perform in the playoffs depending on attendance and payroll. We found that attendance size and payroll are correlated with success, at least visually. Only one team in our dataset had an average attendance under 20,000 and made the World Series.

These results make intuitive sense. Teams that spend more on players usually have more talent, and teams with larger or more engaged fanbases tend to have greater financial and organizational resources. At the same time, the regression makes clear that wins are influenced by many other factors not included in the model. Baseball outcomes depend on roster construction, pitching depth, injuries, random performance swings, and managerial choices. Payroll helps, but it does not fully determine how many games a team wins.

In conclusion, this analysis shows that higher payroll shares are associated with higher win totals, and attendance strengthens that relationship by capturing broader organizational factors. The model gives a clear and consistent answer: **spending more increases a team's expected number of wins**, but a large portion of performance remains outside the scope of payroll alone. Winning in baseball is partly about investment, but also about everything that happens beyond the paycheck.