Absence Prediction & Labor Force Optimization in Rail Dispatcher Scheduling

By Taylor Jensen and Qi Sun
Thesis Advisor: Dr. Tony Craig

Summary: Scheduling dispatchers in the Rail industry is difficult because of training requirements, unplanned employee absences, and the nature of shift work. This project uses past data on absences to build a model to predict unplanned absences. It also uses Monte Carlo simulation to evaluate the tradeoffs among extra staff, the training levels of employees, and required overtime.

KEY INSIGHTS
1. Shift, month, snowstorms, and holidays were all found to influence the number of unplanned absences.
2. There is not a satisfactory way to accurately predict absences on a given day and shift.
3. Increasing the number of employees reduces overtime costs more effectively than increasing the number of qualifications of each employee.
4. The savings in overtime costs from having extra employees does not offset the cost of having extra employees.

Taylor Jensen came to the MIT SCM program after working in Supply Chain Management for NAPA Auto Parts. He has a B.S. in Construction Management from Brigham Young University in Provo, Utah.

Qi Sun came to the MIT SCM program with a B.E. in Computer Science and Technology from the Qingdao University. Prior to graduate school, he worked as a Branch Manager for DHL Global Forwarding.

Introduction
RailCo\(^1\) operates several thousand miles of track across the United States. The department that directs traffic across this network employs over four hundred dispatchers and operates 24 hours a day, 7 days a week, 365 days a year. The daily assignment of these four hundred dispatchers is a difficult problem because every position must be filled with an employee that has been pre-qualified to work in each position. This scheduling problem is further complicated by unplanned employee absences that occur in unpredictable patterns.

Because RailCo cannot predict when days of unusually high absences will occur, they are forced to keep a high number of extra dispatchers on their payroll to cover worst case scenarios. Dispatcher scheduling at RailCo is further complicated by strict union rules that govern employee assignments, qualifications, and overtime pay.

Research Questions
1. What factors affect unplanned absences, and can we use them to accurately predict how many absences will occur on each shift?
2. What are the tradeoffs among the number of extra employees, the number of qualifications of each employee, the amount of overtime required, and total labor costs?

\(^1\) Alias, name has been changed
Methodology

Absenteeism Prediction

In our analysis we used four years of unplanned absences, from 2009 through 2012; these absences were broken down into individual day/shift combinations and placed into a probability distribution. The number of absences that occur on any given shift is categorized as count data because the values are always positive integers.

The most common distribution used to model count data is the Poisson distribution. Figure 1 shows the probability distribution of unplanned absences fit to a Poisson distribution.

![Distribution of Unplanned Absences](image)

**Figure 1: Probability Distribution of Absences**

Knowing that the number of absences that occur on each shift can be modeled with a Poisson distribution allowed us to use Poisson regression to test the influence of a number of factors on absences.

To test the predictive capability of our model we used a regression measure of fit called the McFadden R-squared. This value ranges from 0 to 1, with values close to 1 indicating a high degree of fit.

Simulation & Optimization

To understand the relationships between extra employees, qualifications, overtime, and total labor costs, we used Monte Carlo Simulation (Metropolis, N.; Ulam, S. 1949). In Monte Carlo Simulation, a number of parameters that each obeys a specific probability distribution are defined and used as input into the simulation. By running many iterations of this simulation you can gain an understanding of the best solution to the overall problem. In our case with RailCo, the inputs that obey a probability distribution are the number of absences on a given shift, the number of qualifications of regular employees, and the number of qualifications of extra employees. All three of these distributions could be effectively modeled with a Negative Binomial distribution.

Each dispatcher position must be filled by the regularly scheduled employee, an extra employee, or an employee called from home and paid overtime. We represented the different ways each position could be filled by different employee types with a matrix and used an optimization solver to find the lowest cost solution. The assignments made by the solver are represented by the matrix shown below.

![Assignment Matrix](image)

**Figure 2: Assignment Matrix**

In this matrix each column represents a position that must be filled, and the rows represent employees. A 1 indicates an assignment of an employee to a position. Each position can be filled by a regular employee (rows 1 to N), an extra employee (rows n to N+E), or an employee called from home (row N+E+1). There will be a different cost to fill each position depending on which employee type is assigned. To make the assignments in a way that minimizes cost we used an optimization solver that was designed for this type of assignment problem.

Each iteration of this simulation uses the input from the known probability distributions of absences and qualifications, and the solution from this solver to generate a cost. By running the simulation thousands of times we were able to get an expected
cost given the number of extra employees and qualifications we defined. Then we could change the number of extra employees and number of qualifications, and run the simulation again. In this way we were able to investigate the impact of varying the number of extra board employees and qualifications on costs.

Results

Absenteeism Prediction

We identified several factors that could influence the number of unplanned absences and used Poisson regression analysis to evaluate the statistical significance of each factor. The factors were tested in our regression were

- Day of the week
- Day of the month
- Month
- Shift
- Holidays
- Football Games
- Hunting Season
- Snowstorms
- Planned Absences

The factors that show statistical evidence of increasing the expected number of absences are snowstorms, second and third shifts (when compared to first shift), and the months of January, February, March, April, and December (compared to July, the lowest month). The factors that show statistical evidence of decreasing the number of expected absences are holidays including New Year’s Day, President’s Day, Independence Day, Thanksgiving, Christmas Eve, and Christmas. The factors that do not provide statistical evidence of affecting the number of unplanned absences are day of the month, day of the week, football games, and hunting season.

Despite these findings, the overall predictive capability of our regression model is very weak. The McFadden R-squared value of our model was .018, which suggests that all the factors evaluated to not lead to a good way of accurately predicting unplanned absences.

Labor Force Optimization

The results from our Monte Carlo simulation indicated that overtime costs always decrease when the number of extra employees increases, and the effect is more significant when the number of extra employees is smaller. Increasing the qualification levels of employees (or the number of positions on which each employee can work), was shown to have a very limited effect on decreasing cost.

Another interesting finding is the total labor cost, always increases as the number of extra employees increases. This means that the cost savings in decreasing overtime from extra employees does not offset the cost of additional employees on the payroll. This is illustrated in Figure 3, below.

Conclusions

Using Poisson regression, we identified several factors that provide statistical evidence of influencing the number of unplanned absences. These factors are month, snowstorms, shift, and certain holidays. Despite these findings, the overall predictive capability of our regression model is very weak.

Through Monte Carlo simulation we have shown that increasing the number of qualifications of employees has little effect on overtime costs. Furthermore, overtime costs decrease as the number of extra board employees increases, but the decrease in overtime costs is not enough to offset the increased labor costs of having more employees.