Part 2: Generalized Linear Models

Charles Julien, Chike Odenigbo, Atul Sharma, Gabriel Jobert

11/17/2023

Contents

Instructions

Part 3: Generalized linear models (due November 10 before 11:55 PM)

- Explore various generalized linear models for the response variables of interest, specifically, for the number of rentals (total, AM, and PM). In addition, create a new variable indicating whether the average daily trip duration exceeds 15 minutes, and explore models for this new variable.
- Be sure that your analyses allow you to answer well formulated business / research questions that you wish to address. The goal is to use generalized linear models to provide interesting and relevant insights from the data.
- Comment on findings and discuss the main takeaways from this analysis from a business perspective. Be sure to provide relevant model outputs that support your discussion.
- Discuss any shortcomings or limitations of the analyses carried out.

Introduction

Business/Research questions

The target variables is number of rentals (total, AM, and PM)

Pre-processing

Imputation (might have to remove)

Revenue for members is missing since they do not pay a usage fee, but rather a fixed cost.

```
imputation_model <- lm(rev ~ dur + avg + n_tot , data = df_main)
df_main$rev_pred = predict(imputation_model, df_main)
```

To impute revenue for members, we make the assumption that they would bring in as much revenue as non-members for the same usage.Thus we consider the same formula of revenue used for non-members. This unknown deterministic function is most likely a linear combination of usage variables like `dur`, `avg` and `n_tot`. We try to approximate this function and use it to impute members revenue. The imputation model has an r-squared of 1 on non-members data.

Research Question 1: How does … ?

Objective of Analysis:

Variables Selection

Model

Interpretation

Overall Model: R squared, f stat

Intercept: The intercept is …

Business implications (can change the sub categories)

1. **Operational Adjustments:**
2. **Rainy Day Strategies:**
3. **Membership:**

Research Question 2: How do. . . ?

Objective of Analysis:

Variables Selection

Model

Interpretation

Overall Model : R squared F stat

Intercept*:

Business implications (can change the sub categories)

1. **Operational Adjustments:**
2. **Rainy Day Strategies:**
3. **Membership:**

Research Question 3: What variables . . . ?

Objective of Analysis:

Variables Selection

Correlation:

Let's take a quick look at the correlation between our numerical variables to estimate the effect of collinearity.

```
##                   avg        temp        rain        n_tot   percent_AM
## avg         1.00000000  0.09639054 -0.10619900 -0.215866274 -0.107387372
## temp        0.09639054  1.00000000 -0.02794911  0.139997362 -0.078110564
## rain       -0.10619900 -0.02794911  1.00000000 -0.054717667  0.013211523
## n_tot      -0.21586627  0.13999736 -0.05471767  1.000000000 -0.008953075
## percent_AM -0.10738737 -0.07811056  0.01321152 -0.008953075  1.000000000
```

Model

Interpretation

Overall Model R squared, F-stat

Intercept :

Business Implications:

1. **Promotion and Marketing**:
2. **Resource Allocation**:
3. **Pricing Strategy**:

Limitations and shortcomings

Autocorrelation of data, the observations are not independant, as seens in our previous analysis.

Conclusion

Contribution

Charles Julien :

Gabriel Jobert :

Chike Odenigbo :

Atul Sharma :