STAT 306: Group Project Final Report

Introduction

Data Collected

Our Question / Motivation behind study

<can probably copy a lot of this from our proposal>

Analysis

EDA Stuff

Scatterplot analysis don’t tell us much.

Correlations tell us humidity, temp and atemp have the strongest correlation (among variables that could influence number of riders)

Temp and atemp have 0.98 correlation (as expected really), so doesn’t make sense to use both of them in model.

Weekends are not marked as either holiday or workingday.

Weather variable has;

|  |  |
| --- | --- |
| Factor | Observations |
| 1 | 7192 |
| 2 | 2834 |
| 3 | 859 |
| 4 | 1 |
| Total | 10886 |

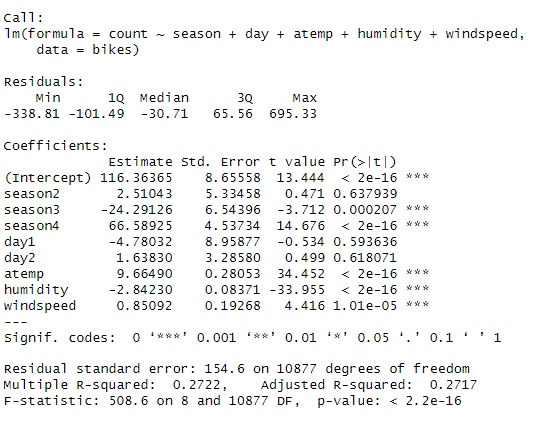
As 66% of the observations are marked as best weather, might not have useful information for predicting number of bikes rented.

Weekends are marked as workday=0, holiday=0. So we’ve made a new variable day where day = 0 means workingday. Day=1 means holiday. Day=2 means weekend.

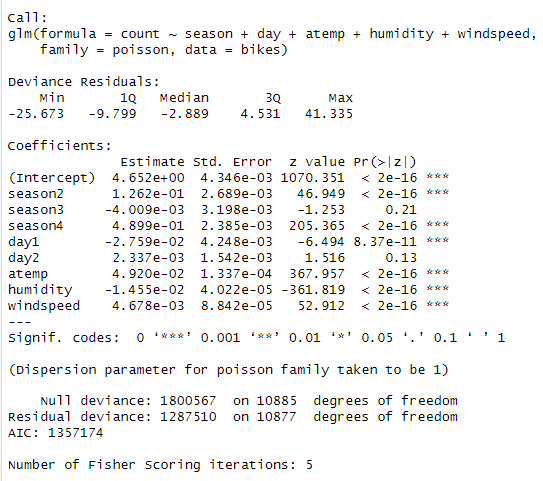
Models

Basic Additive model;

Chose to first setup the following model



Counting data so poisson regression makes more sense



AIC much lower when using Poisson Regression.

Conclusion

Continuous data always significant across all models tested.