**How long will you wait for your coffee?**

ECON 124 Machine Learning

Final Paper

by

Liam Jay, Zack Renick, and Terra Kelly

1. **Introduction**

Coffee shops have become an integral part of modern-day society, offering a convenient and quick source of caffeine for millions of people worldwide. However, long wait times can often deter customers from returning to a coffee shop, leading to a loss in business. As such, predicting wait times accurately can greatly benefit coffee shops by allowing them to optimize their operations and minimize customer wait times.

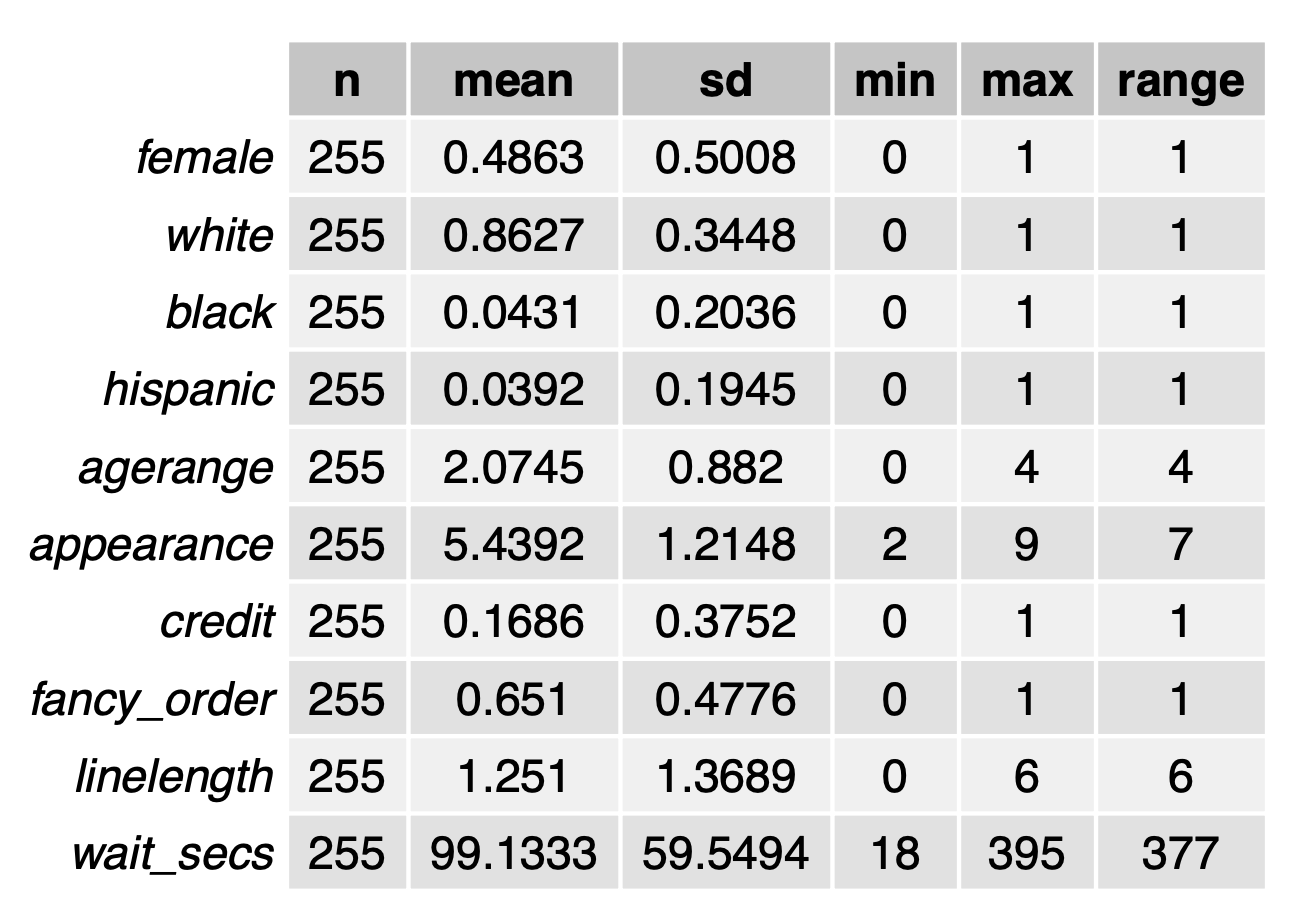
In this paper, we utilize a dataset containing information about individuals' characteristics, coffee shop locations, and the duration of their wait time to build a predictive model using Lasso regression. Our objective is to develop a model that accurately predicts wait times based on individual characteristics, such as age, gender, race, and the type of coffee order placed. We aim to provide insight into the factors contributing to long wait times, enabling coffee shops to improve customer service and optimize their operations.

Our research is significant because it addresses a common problem coffee shops and their customers face worldwide. Accurately predicting wait times can help coffee shops to reduce customer wait times, improve their customer experience, and ultimately increase revenue. Additionally, the use of machine learning models, such as Lasso regression, can enable us to identify the most significant predictors of wait times, providing further insight into the factors contributing to long wait times.

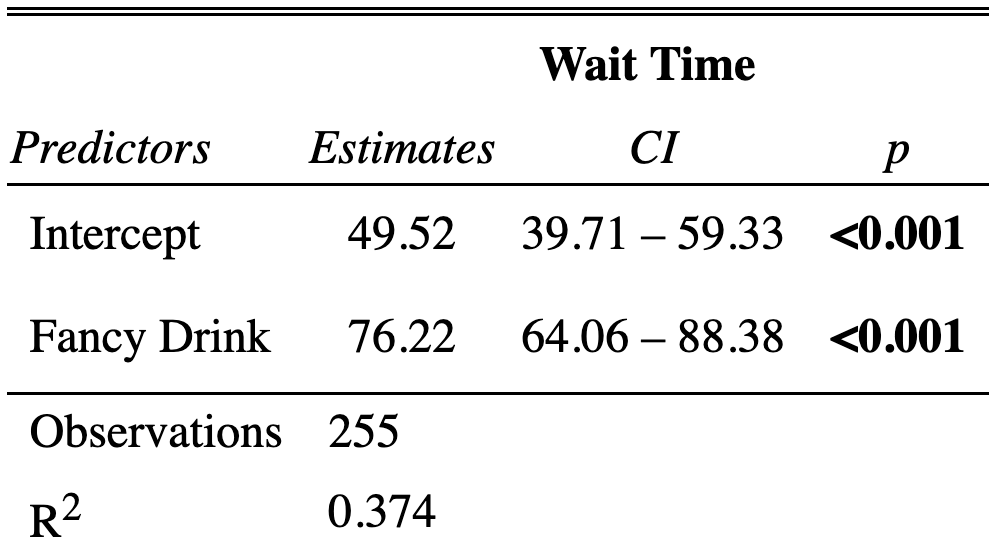
Our paper is organized as follows. We will begin by providing a detailed description of the dataset used in our analysis. Then we will outline the methodology used to build the Lasso regression model. Next, we will present the results of our analysis, including the model performance and significant predictors of wait time. Finally, we will conclude our discussion by summarizing our findings and discussing the implications of our research for coffee shops and the wider community.

1. **Data**

The data for our analysis comes from a 2008 study by Caitlin Knowles Myers that looks at gender discrimination in coffee shops, specifically, whether there is a difference in wait times for female customers to get their drinks compared to men. The researcher gathered data from eight coffee shops across the greater Boston area and recorded data such as a customer's gender, age range, and ethnicity, as well as the time of day, line length, and what type of payment was used. As our analysis hopes to compare which ML method best predicts a person's wait time, we will utilize all available covariates to maximize predictive power. The regularization aspects of Lasso and the Glmnet penalization will narrow the chosen covariates to only the most effective predictors.



One main takeaway from our dataset is that it is heavily skewed white. Over 85% of the observations are of people the enumerators described as white. This is something to consider when we think about the representativeness of our sample as a proxy for the entire United States. It's also worth noting that in the sample, only 16.86% of customers paid using a credit card. While this study was done more than a decade ago, this also would not be the most representative for coffee shop patrons in today's world where most customers pay with a mobile payment or a credit card. Neither of these things invalidate our data or the prediction we are trying to make. Rather they are things we must consider in regards to the rigor of our predictions. As a data scientist for Starbucks, we would not want to use this naive model as *the* prediction tool when trying to understand how long it will take for customers to get their drink because we know that there are flaws in the representativeness of the data from which it was created.



One initial result that we found was that people that ordered a fancy drink (ie. latte, mocha, etc) often ended up waiting a much longer time than those that did not. While this could be expected, we must also consider the fact that a person's drink order may be correlated with other characteristics about a person.

1. **Methodology**

After cleaning the dataset, we will use Lasso regression as the primary modeling technique to predict the outcome variable of wait time in seconds. To ensure the accuracy of our model, we will use 10-fold cross-validation, which involves dividing our data into 10 equal parts, training the model on 9 parts, and validating the results on the remaining part. We will repeat this process 10 times, ensuring that each part of the dataset is used for validation once.

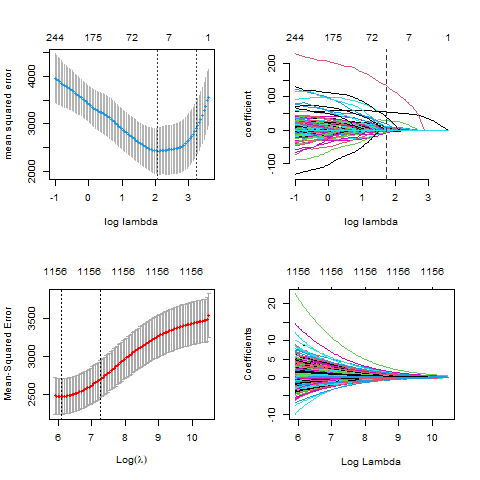
We will also perform a training and test split of our data, with 80% of the data in the training set and the remaining 20% in the test set. By doing this, we can evaluate the performance of our model outside the model-building process. We will compare the in-sample (IS) and out-of-sample (OOS) deviance of our Lasso model to establish its predictive power. The deviance is a measure of the model's goodness of fit and represents the difference between the predicted probabilities and the actual outcomes.

In addition to Lasso regression, we will also investigate how Ordinary Least Squares (OLS) and Ridge regression models perform on the same dataset. This is because our dataset has low dimensionality, and we may only want to remove certain covariates partially. Ridge regression is another type of linear regression that adds a penalty term to the model to reduce the magnitude of the coefficients. OLS regression, on the other hand, is a standard linear regression that aims to minimize the sum of squared errors between the predicted and actual values.

By comparing the performance of these three models, we can identify the most appropriate model for predicting wait times based on individual characteristics. Ultimately, the objective of our analysis is to identify the most significant predictors of wait times, enabling coffee shops to optimize their operations and reduce customer wait times.

1. **Main Results**

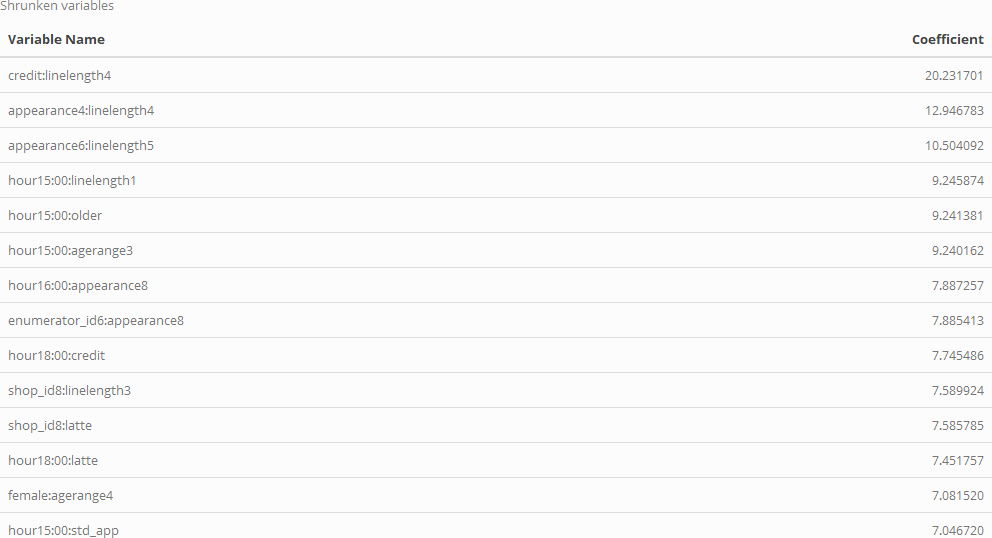
As it is our goal is to predict wait times for a person at a coffee shop until they receive their order we must come up with some way to model the data in order to predict on. The simplest approach would be to fit an OLS model on the data; this will be the first method we explore. We could naively fit a “kitchen sink regression” (regressing wait time on all the covariates and their interactions with each other) for all of our data and then predict from that, but how accurate would this model prediction be? We can check this by doing a 90-10 sample split with our coffee data. By allocating 90% of the data from coffee into a “training set” and the remaining data into a “test set” we are able to train our model on a large portion of the data and then evaluate how well the predictions obtained using said model perform through computation of the OOS deviance to obtain the OOS R^2. In this case our “test set” is serving as the proxy for some new data our employer at Starbucks may supply us with and would be interested in knowing wait time predictions for. For this naive model we observe an OOS R^2 value of -322598926; this is a really poor result. Our OOS R^2 is extremely large in magnitude as well as negative. This means our model is predicting way worse than even the null model with no covariates would be. To correct for this we can use some regularization penalties such as Lasso and Ridge to improve the selection and shrinkage of our coefficients in the hopes of improving the predictive power of our model. These regularization methods were both estimated using 10 fold cross validation for model selection and the regression specifications were kept the same as with OLS. We find that the optimal lambda penalty for Lasso is 7.827088 and the optimal segment is 34, while the optimal lambda penalty for Ridge is 437.5977. Included below are the error-lambda and regularization plots for the Lasso and Ridge models respectively; with Lasso being the top row and Ridge contained in the bottom row.

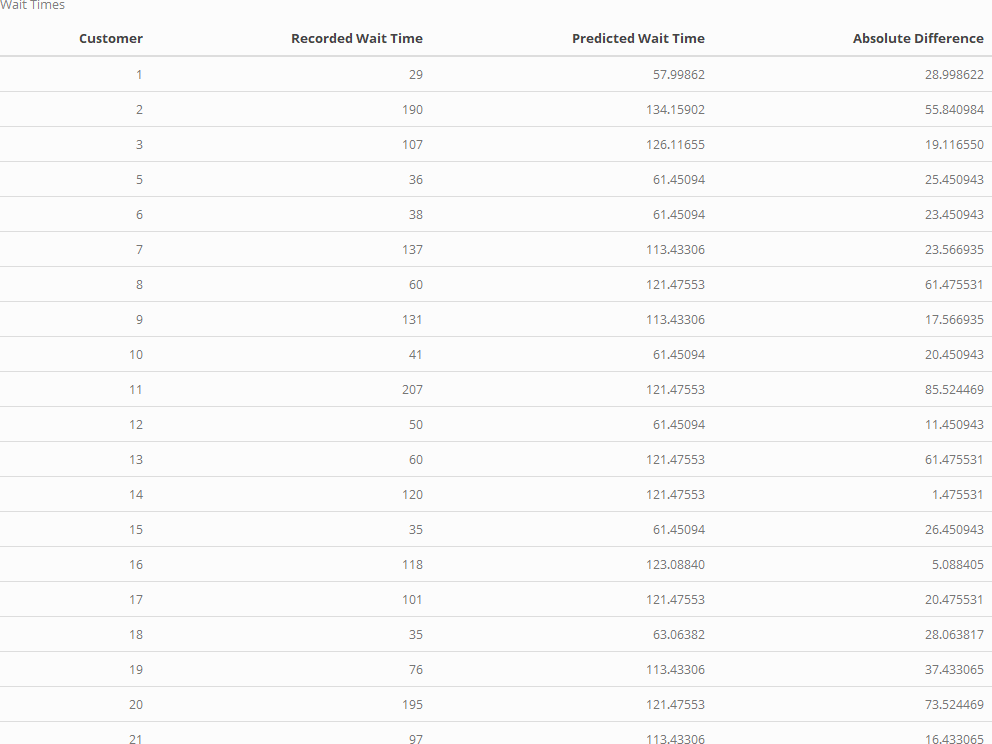


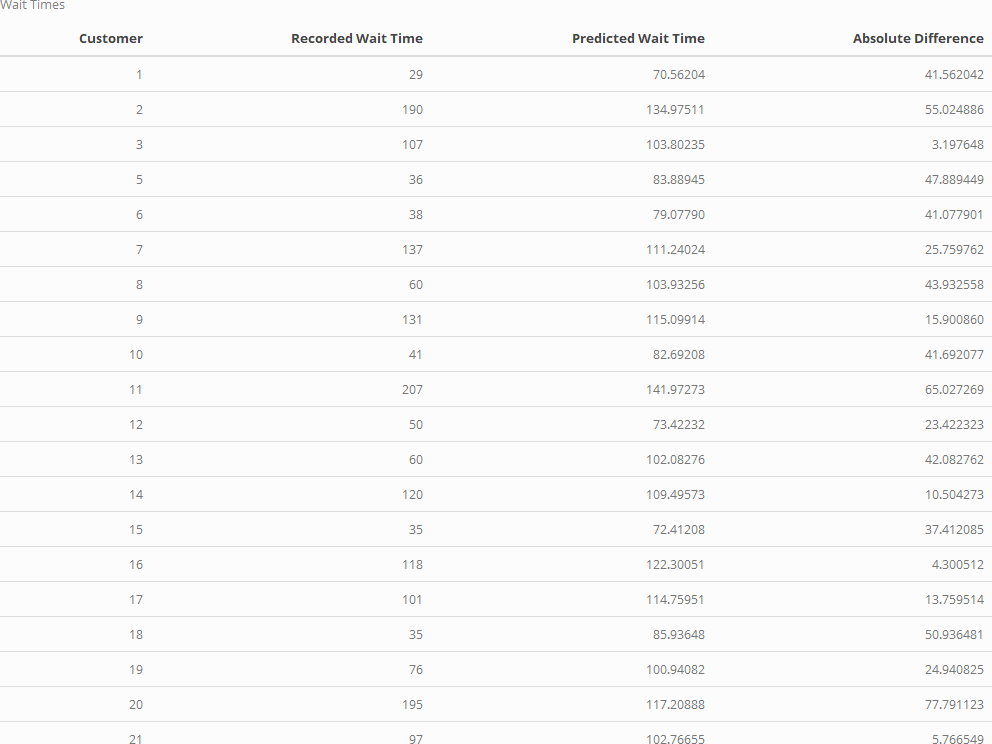
So which coefficients did our regularization models deem the most important and assign the most weight too? Recalling that the lasso penalty will zero out coefficients during regularization and Ridge will typically shrink coefficients towards zero we find that there are 14 nonzero coefficients in our Lasso model.



We can compare these coefficients to the 14 coefficients Ridge determined to have the most predictive power.



Across the two models we see they placed a large amount of weight on the line length as we see those coefficients interacted with others frequently across the tables. We also observe that while both models do prioritize characteristics about the way someone looks (appearance, gender, age, and race) the Lasso model finds more predictive power using coefficients about the coffee order itself. As opposed to Ridge which is placing a lot of emphasis on which shop a person visited and at what time. In this case the regressors included and excluded are about what we would expect for an order wait time. A lot of this time depends on simply how long you have to wait in line in combination with how complicated your order was. Along with some other wait time increases accounted for by either subconscious or intentional bias of the baristas based on human characteristics which we know is found across industries. The next obvious question to answer is how well did this predictive model actually perform? First, we present a table summarizing the first 20 observations to get a flavor of the performance for each model, Lasso and Ridge Respectively.



It is clear even from these tables that our model does not have very high predictive power since we see at least double digit second differences for almost every customer across models. We can, however, quantify this performance by once again using the OOS deviance of the models to compute the OOS R^2 and determine the model fit to the data. We find the Out Of Sample R-squared using 10-fold cross validation for Lasso as 0.3163004 and 0.3016375 for Ridge. While this result is certainly not great it is significant better than the simple OLS model run previously and more specifically if we were to present this to an employer at Starbucks the Lasso model would be chosen for its slightly better predictive performance out of sample as compared to Ridge. While an employer may find some use out of this model to estimate wait times, its lack of predictive performance is potentially concerning for actual industry use.