Jielin Cao

March 3, 2020

How Much Do Consumer Tip in Restaurant?

Abstract

Tipping, like any other social skill, is a learned behavior, taught either by a person from an older generation or through life experience (Spector). Tipping has become a powerful ritual in United States, it becomes so common that, it influences people's live styles. According to Michael Lynn, a researcher in Cornell University who has been studying tips for years, on an average day there are approximately thirty million people in united states eat at restaurants that have servers, and on an average month, the number of people eat at restaurant reaches a hundred fifty million. These tips, which amount to approximately $21 billion a year, are an important source of income for the nation's two million waiters and waitresses (Lynn). Tips sometimes represent 100 percent of server take home pay when tax withholding has cost all the hourly wages (Lynn). According to TIME, tip has fully subsidized a bartender or server's salary at the vast majority of the nearly 650,000 restaurants in the country (Greenspan). Especially when the minimum wage is not enough to pay off the living expenses in northern California, tip becomes precedentially important.

Introduction

Tipping tradition was discovered by "…[w]ealthy Americans in the 1850s and 1860s, which had originated in medieval times as a master-serf custom wherein a servant would receive extra money for having performed superbly well, on vacations in Europe" (Greenspan), from then on, tipping has started its prevalence. Although tipping is a random behavior, there are still patterns for us to analysis. For instance, as consumer, we always know the appropriate bound of tip. According to NBC News, tipping has evolved into a social norm, and every one of us knows the rules. For bartender, we tip a dollar per drink, for server, we tip 10-20%, for hotel housekeeping, we tip two dollars per night, etc. Besides tipping in restaurant, tipping in bar, in hotel, in taxi, in salon, in parking lot are also the cases (Spector).

The main purpose of this paper is to predict the amount of tip under different scenarios and circumstances, in what way consumer is willing to tip more generously? As I mentioned above, tipping is powerful in United States, many people live on it, and sometimes in lieu of the hourly wages. I believe that this paper will be of great interest to the people who work in service sector, because this paper is dedicated to predicting the amount of tip customer gives to the server using a few interesting variables such as, the total bill amount, the customer's gender, whether there is a smoker in the party, the day, the time, and the party size. Among these variables, I am trying to determine the most influential one adopting lasso method and cross-validation method.

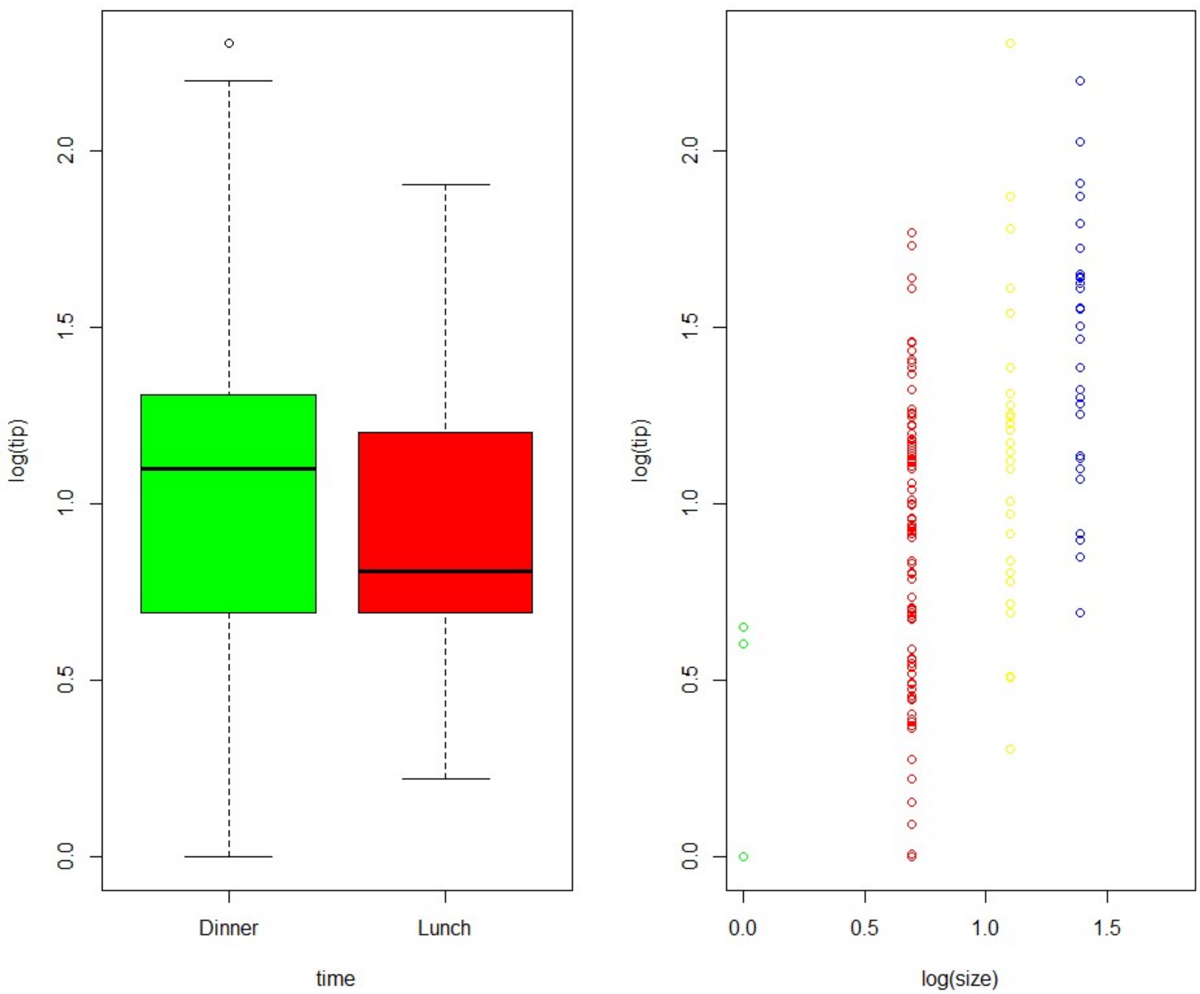
Data

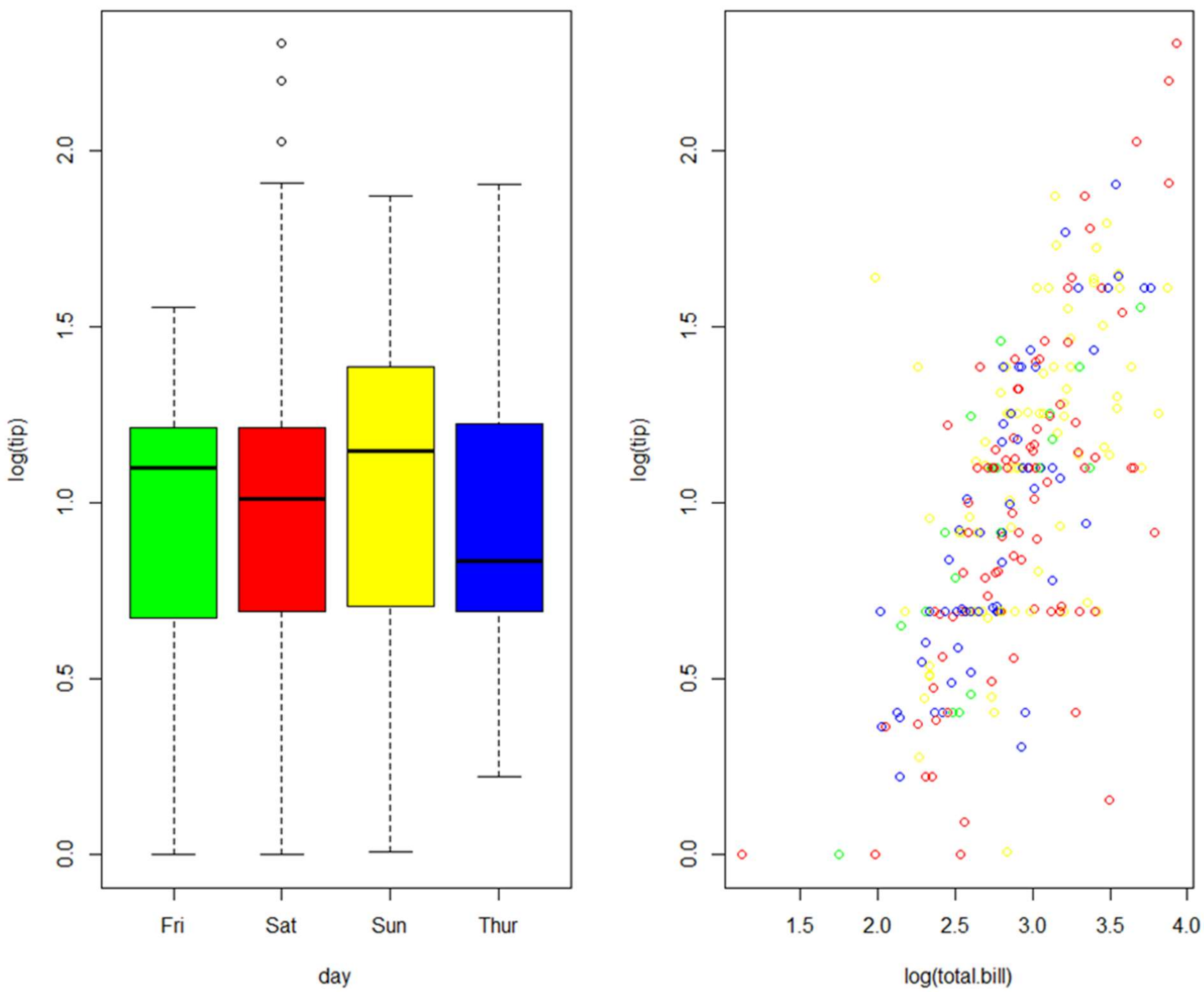
I used a dataset from Kaggle, an online community of data scientists and machine learning practitioner subsidized by Google LLC (Wikipedia). The dataset is originally collected by a waiter while working in a restaurant. When receiving the tip, the waiter carefully recorded all the information. There are 244 tips recorded in total, which means there are 244 observations in the dataset. According to the waiter's description, it took a few months to collect all the information. Also, this dataset is included in *Practical Data Analysis: Case Studies in Business Statistics* by Bryant, P. G. and Smith, M (1995).

First, by accessing the dataset, I found out that the average amount people tip is $2.99 with a standard deviation of $1.38. In estimating the population mean, the standard error is 0.089, which leads us to the confidence interval between $2.82 and $3.17. The difference between the sample average and the population mean here is small. In order to find out how plausible these numbers are, I run a bootstrap trying to estimate the population mean. I found that under bootstrap, the standard error is 0.087, which is very close to the sample standard error above. In the sense of using the sample to estimate the population mean, undoubtedly, this dataset did a good job. Now, bootstrap interval is between $2.83 and $3.17, which is still closed to our sample interval, seemingly, the numbers are plausible.

On average, male tips 25 cents more than female does, and smoker tips 1 cent more than non-smoker does, which is no different to me. From another perspectives, based on the elasticity of tip, smoker would be likely to increase the amount of tip by 0.06%, and for male is 0.03%. These numbers seem insignificant, yet, surprisingly, female smoker pay higher amount of tip than non-smoker, but for male, situation is opposite, male smoker tip less. In general, male tip heavier than female. In this case, if I was the server, female smoker and male non-smoker are my best bet when I am trying to collect the gratuity. Also, the dataset shows the effect of day, time, and the total bill amount. For example, Sunday, compared to the other three day, which are Thursday, Friday, and Saturday, has the highest tip average, which is $3.26. Dinner has $3.1 tip average compared to \$2.73 for lunch, If I were working as a part-time server, I would register myself to the Sunday night shift. For the total amount of the bill, and the size of the party, they both clearly shows the positive relationship with amount of tip. The average amount of tip collected is \$5.23, when the party is six, \$4.03 when it is five, \$4.14 when it is four, \$3.39 when it is three, \$2.58 when it is two, and \$1.44 when there is only one. It is the same for the total amount of bill.

In terms of elasticity between the dependent variable (tip_i) and all the independent variables, I plotted four graphs. The first one is the elasticity between ($time_i$), ($size_i$) and ($log(tip_i)$), as you can clearly see that percentage changes more at dinner time, above 1% compared to below 1% for lunch, and it changes more for bigger party. Next graph is the elasticity between (day_i), ($total.bill_i$), and ($log(tip_i)$). Also, we can see that Sunday has the biggest elasticity among days, and for the graph on the right, we can also see the positive relationship between ($total.bill_i$), and ($log(tip_i)$).


log(tip)
0.0
0.5
1.0
1.5
2.0
Dinner
Lunch
time
log(tip)
0.0
0.5
1.0
1.5
2.0
0.0
0.5
1.0
1.5
log(size)



Model

In determining the effect on the amount of tip customer gives, an econometrics method is adopted. For every potential outcome (*tip_i*), there is a corresponding factors (*i*), which determines the characteristic of the tip payer, the total amount of the meal (*$total.bill_i$*), whether the person smoker (*$smoker_i$*), the gender of the person (*sex_i*), the day when the tip was collected (*day_i*), the time when the tip was collected (*$time_i$*), and the size of the party (*$size_i$*). In terms of the ability upon prediction, I am going to use cross validation method, since we need to train the model to do its best at predicting the amount of tip. By doing so, I divided the observations into 20 different folds, randomly pick 19 of them to train my model to predict left out fold, then I record the average deviances and I will do it twenty times in order to pick the one with the lowest deviances. Also, I will use LASSO method,

$$\hat{\beta} = \min_{\beta} arg \sum_{i=1}^{n} (y_i - x_i'\beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

where tuning parameter λ will be chosen by cross validation for out-of-sample deviance to sort out some of the most effective variables by penalizing the regression to the most effective extent. From the first equation, which is the most common regression equation.

$$tip_i = \beta_0 + \beta_1 total.bill_i + \beta_2 sex_i + \beta_3 smoker_i + \beta_4 size_i + \beta_4 time_i + \beta_5 day_i + \epsilon_i$$

I found that there are only two variables are significant, which are total bill amount and size of the party, in other words, only these two variables have p-values that are lower than 0.05. In order to discern heterogeneity, I slightly modify the equation by adding a few more interaction terms,

$$\begin{aligned} tip_i = \beta_0 + \beta_1 total.bill_i + \beta_2 sex_i + \beta_3 smoker_i + \beta_4 size_i \\ + \beta_4 time_i + \beta_5 day_i + \beta_6 smoker_i * sex_i + \beta_7 total.bll_i * size_i + \beta_8 total.bill_i \\ * time_i + \beta_9 total.bill_i * day_i + \beta_{10} size_i * time_i + \beta_{11} size_i * day_i + \epsilon_i \end{aligned}$$

so that the impact of relationships among variables would be eliminated to a great extent. As we have already mentioned above, there are only two variables are significant, I only include the interaction terms that are related to them.

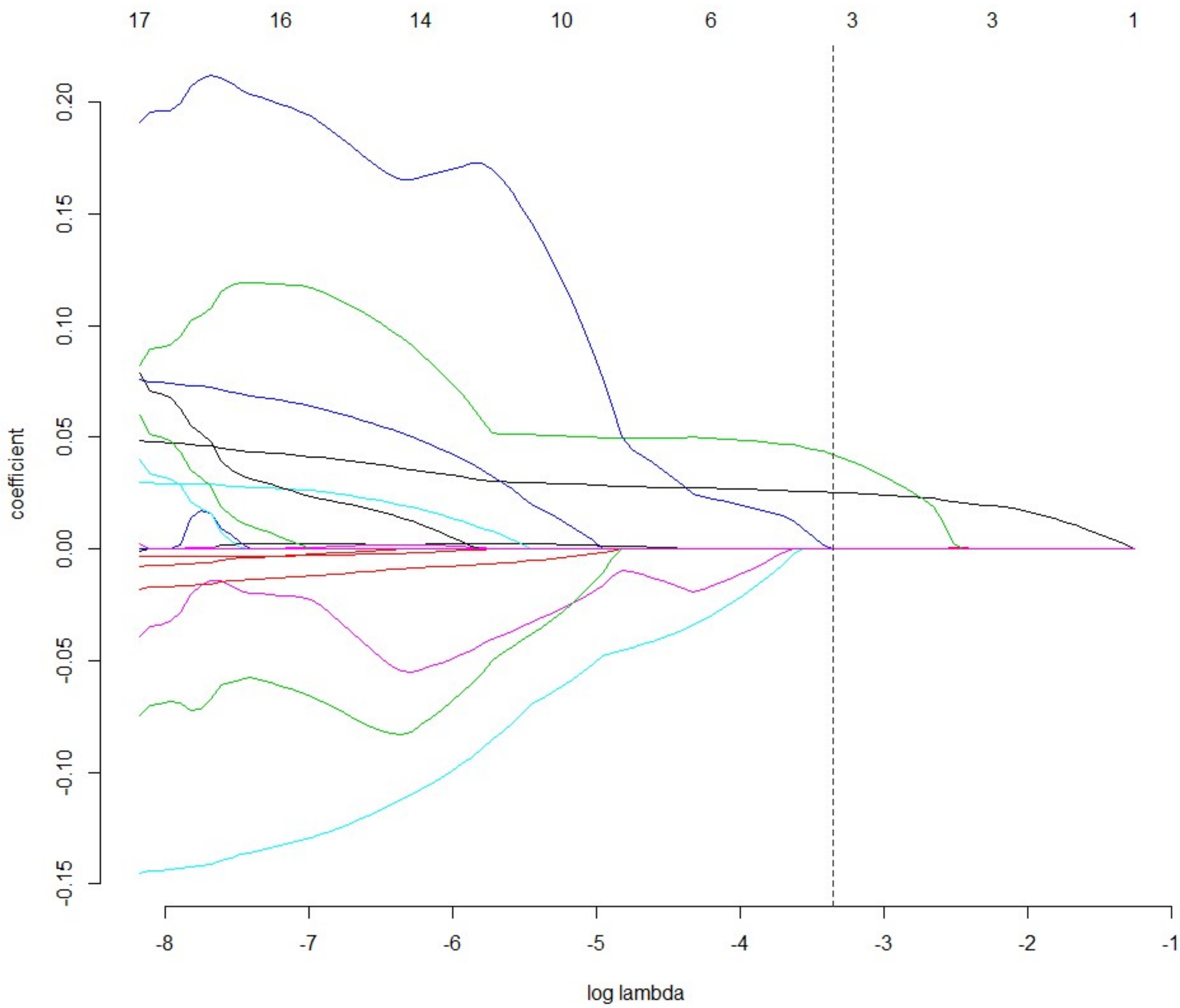
The second equation is trying to represent all the potential cases, and to avoid collinearity, in which variables are possibly affecting each other, resulting in an off-track prediction. For instance, the term ($smoker_i$ * sex_i) intended to eliminate the effect of smoker on sex, since majority of the smoker are male. Also, the total amount of the bill attributes to lots of other factors, for example, dinner is usually more expensive than lunch, so I added the term ($total.bill_i$ * $time_i$). Besides, the bigger the party, the bigger the bill, and there is ($total.bill_i$ * $size_i$). Moreover, after taking those affecting the size of party into consideration, such as day and time, because people are more likely to gather when it is weekend, or when it is dinner time, so there are ($size_i$ * day_i), and ($size_i$ * $time_i$).

Result

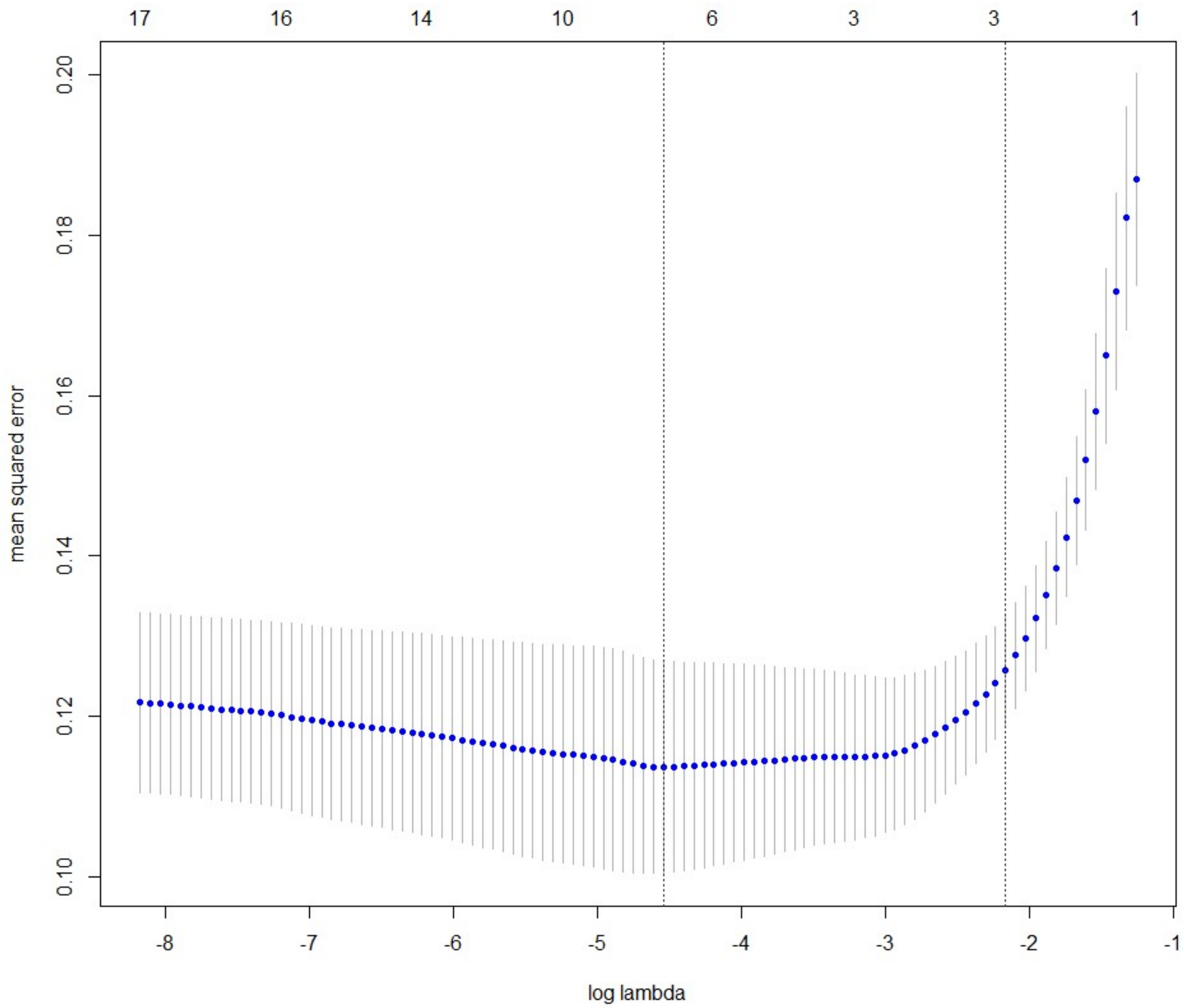
Estimation results shows that in-sample deviation for the first equation is 0.47, meaning the dataset captures 47% of the observed variation. For the second one, in-sample deviance increased to 0.503, meaning now, the improved model is able to capture 50.3% of the observed variation, which is better, even though increasing the number of regressor generally increases the in-sample deviance. Now, let's look at out-of-sample deviance, which tells us how well the model does in predicting the population amount of tip. According to the model, out-of-sample deviance from the first equation is -0.3792, meaning the fitted model managed to do 37.92% worse than the null. For the second one, out-of-sample deviance is -0.3646, we could see that there is an enhancement yet, it is still closed to the first equation. In other words, our fitted model is even worse off predicting the population tip amount, compared to the overall tip average in the dataset. According to the textbook, along with the cross validation for "full" model, there is also a "cut" one, which means we manually select useful variables by sorting out the regression that have the highest p-value frequency. I did that as well, for the first equation, I obtained -0.3851, which is worse, but for the second equation, I got -0.2981, which could be considered a small improvement.

In terms of coefficients, they represent the effect of each variables have upon our dependent variable, which is the amount of tip. With a p-value equals to 0.0289, the coefficient of ($total.bill_i$) is 0.1, meaning a dollar change in total bill amount is associated with a 10 cents change in tip, and it is statistically significant. The coefficient makes sense because that is the tipping percentage people usually apply on their bills, it bounces between 10-20%, and it also depends on the type of the restaurant. Besides ($total.bill_i$), (sex_i) and ($smoker_i$) also play important roles in this dataset. Accordingly, Male customer would generally tip 15 cents more than female customer does, and smoker would tip 21 cents more than non-smoker. Although the two variables have a surprisingly high p-value, which are 0.401 and 0.378 respectfully, they do make sense as the tip average for male and smoker is higher.

As there are so many disqualified variables in the dataset, next step, I am going use lasso method to get the optimal complexity of my model using the second equation and, combined with cross-validation method, we can go through the model selection process. The graph below depicts the lasso path plot for my model. The vertical slice here represents the candidate model. For the graph below, going from the right to left, the model is getting complex, since λ is getting closed to zero, which mean the coefficients are not being penalized at all (non-zero). It would be the opposite if we are going from left to right. The vertical dashed line is the most recommended value for λ, which corresponds to the number ranges from “6” to “3” at the top of the graph.



For the graph below, it depicts the out-of-sample prediction performance for each λ for my model. Again, going from right to left, the model becomes more complex as the value of λ is getting closed to zero. The two vertical dashed line denotes the CV min rule, and the CV 1se rule, corresponding to the somewhere between number "10" and "6", and somewhere closed to number "3" at the top of the graph. The CV min rule tells me the λ value when out-of-sample deviance is minimized, and CV 1se rule tells me the λ value when deviance is one standard deviation away. In summary, it selects 6 out of 12 variables in my model. In summary, (*total.bill*$_i$) is the parameter of interest as it is closely related to the amount of tip in my project.



Conclusion

Analyzing dataset using bootstrap, cross-validation and lasso method can be delightfully helpful when we are trying to interpret and understand the dataset and select the best model for predicting. In this project, I used three of those method to analyzing, sorting, selecting my final model to predict the amount of tip customer give in a restaurant. My analysis shows that the total amount of bill is positively related to the amount of tip customer gives out. Each dollar changes in total bill amount is associated with 10 cents change in tip amount, which makes sense because that is the percentage people usually tip their server. In general, servers collect more tip during weekend than weekdays, especially dinner, and they collect more from customer who are male and smoke. Besides, size of the party also plays an important role, every person increased in the party results in 18 cents increase in tip amount. Yet, overall, the dataset does not do well in prediction, although in-sample deviance reaches 50%, meaning it is able to capture half of the sample variation, out-of-sample deviance is negative, which means its prediction is worse than the sample tip average.

It has been interesting to study the motivations and significance behind tipping, not only because it has become a norm and even a measurement, but also because it is full of randomness. Some people consider tip as miscellaneous expense or pocket money and tip for totally nonsensical reason (Tuttle), yet some live on it. Some industries waive tip for customers, yet some are trying to signal customer for bigger tip (Elliott). According to Micheal Lynn, many economists consider tipping as “mysterious” or “seemingly irrational”. In my opinion, if we have sufficient dataset and patience, we can turn “mysterious” into “known”.

Reference

Elliott, Christopher. “Is Tipping on the Way out? Here's Why More Travelers Are Joining the 'Do Not Tip' Movement.” *USA Today*, Gannett Satellite Information Network, 17 Feb. 2020, www.usatoday.com/story/travel/advice/2020/02/14/tipping-while-traveling-why-gratuities-may-going-out-style/4748441002/.

Greenspan, Rachel E. “How Americans Tip at Restaurants Has a Troubling History.” *Time*, Time, 20 Aug. 2019, time.com/5404475/history-tipping-american-restaurants-civil-war/.

Lynn, Michael. “The Effects of Tipping on Consumers Satisfaction with Restaurants.” *Journal of Consumer Affairs*, vol. 52, no. 3, 2017, pp. 746–755., doi:10.1111/joca.12171.

Spector, Nicole. “Dollars and Sense: Why Are Millennials Tipping Less than Older Generations?” *NBCNews.com*, NBCUniversal News Group, 27 June 2018, www.nbcnews.com/better/business/dollars-sense-why-are-millennials-tipping-less-older-generations-ncna886966.

Tuttle, Brad. “15 Things You Didn't Know About Tipping.” *Money*, 19 Sept. 2014, money.com/tipping-myths-realities-history/.