Do Men Tip More Than Females?

Natasha Leodjaja

May 29, 2022

Abstract

Do you tip when you go eat in restaurants? If so, how do you determine the amount of tip to give? Does size party influences the tip amount? Bigger parties tip more? Do females tip more than men? A lot of people such as Economists have a hard time to solve this issue or explain this problem. This study investigates factors that influence tip amounts in the service industry such as gender, party size and day of the week. It was first tested by a two sample t-test to compare the means of tip given by male and females. Then I build a model using Multivariate linear model, the model can make decent predictions about the response, which can be used to predict tip amount and make rough speculations about tipping in real life.

Introduction

Growing up in Indonesia, I have never once tipped our food servers and I was curious to see how much do waiters/waitress in the US earn by serving a party. In the US, tipping is an understood necessity of the service industry to balance out the standard server wage of approximately $2.13 an hour. With so many individuals working in the service industry and making a living, it is estimated that billions of dollars are spent on tipping each year (Nelson 2017). The dataset used for this study was collected by the food server of a restaurant. The goal of this study or analysis is to evaluate whether certain factors such as gender, day of the week and party size contribute in determining the amount of tip that food servers get.

In this study, we will be evaluating said factors and use the tips dataset to test out two hypotheses. I hypothesize that the average tip given by men is equal to the average tip given by females. I also hypothesize that days, gender and group size are not a significant predictor of tip that waitress/waiters earn. This study attempted to build a multivariate linear model which allows us to see which variables affect the tip amount the most.

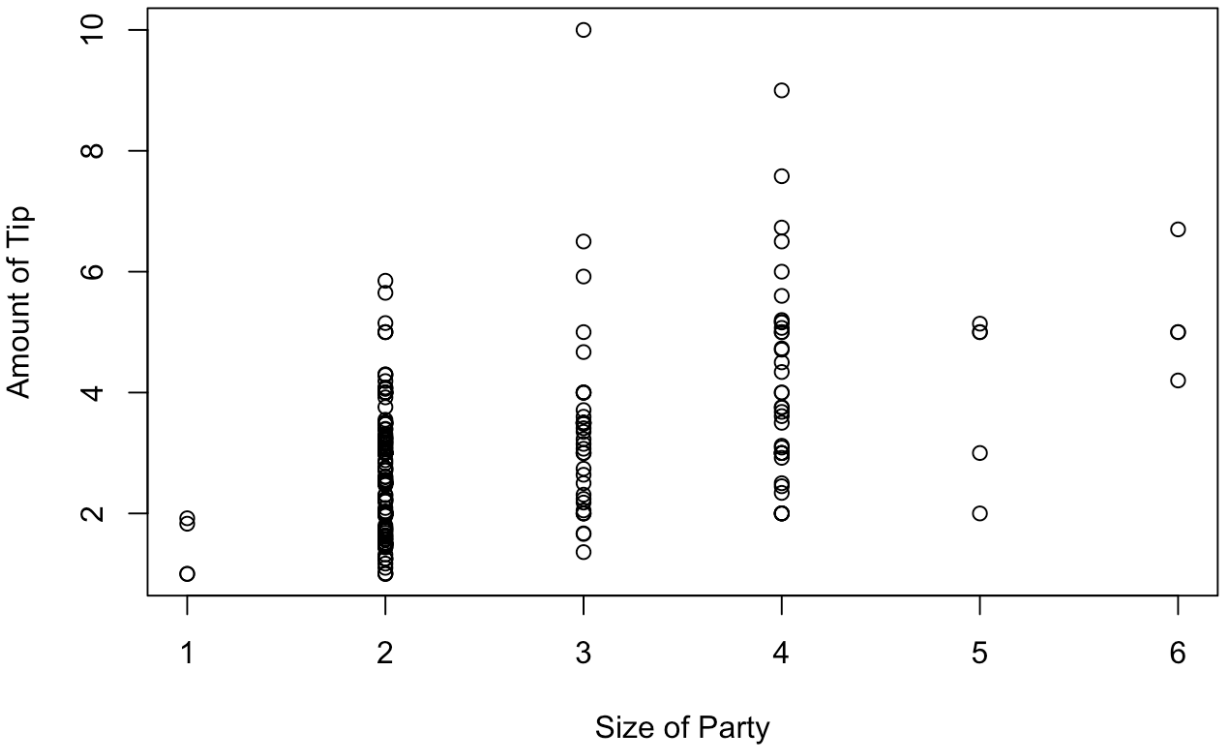
There's about 244 rows in this dataset with 7 variables. However, in this study, we will only be focusing on 4 variables that is tip, sex, day and size. Where tip is our outcome of interest and the rest are our inputs of interests.

Tips Dataset

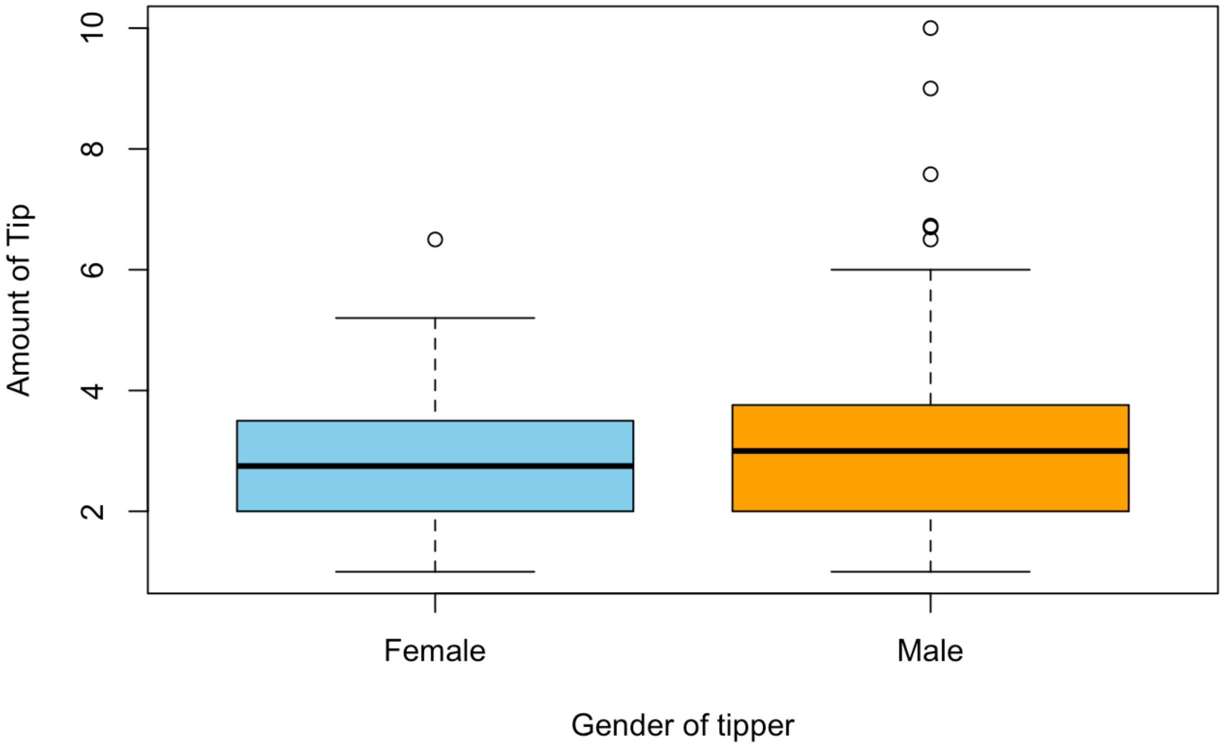
sex	day	size	log_tip
Female	Sun	2	0.0099503
Male	Sun	3	0.5068176
Male	Sun	3	1.2527630
Male	Sun	2	1.1969482
Female	Sun	4	1.2837078

Exploratory Data Analysis

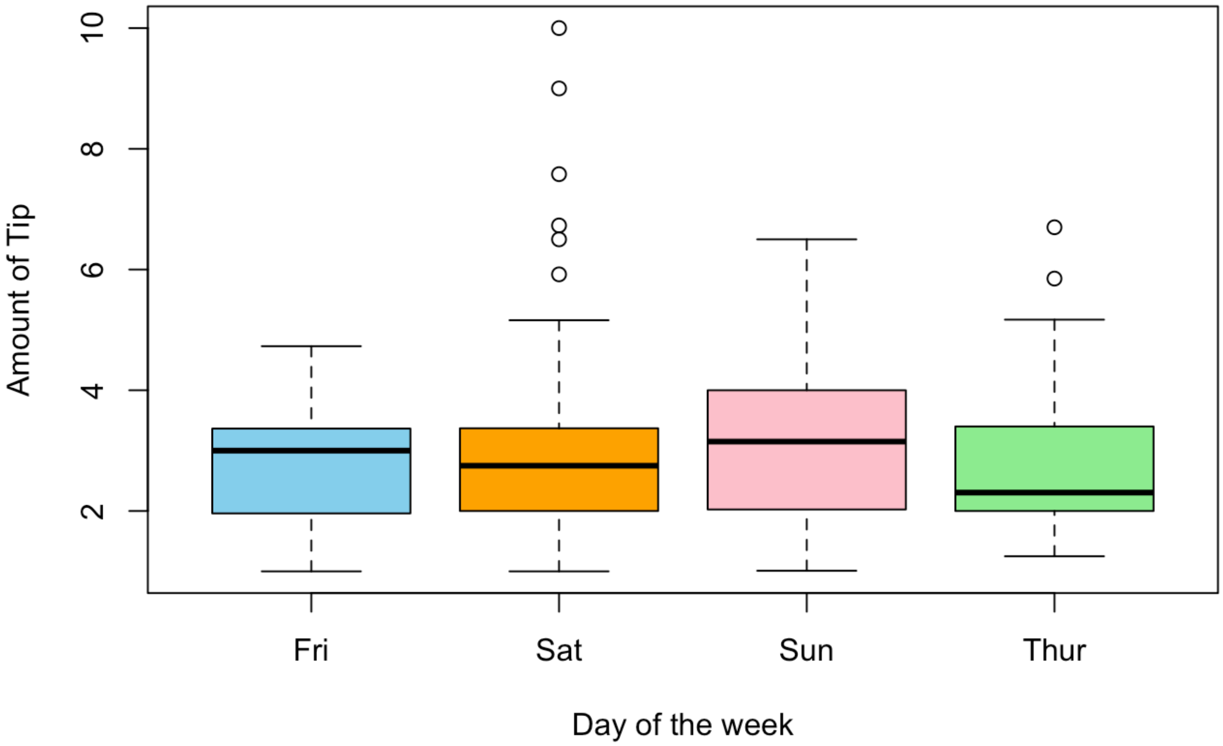
There is not much to do for data cleaning as the dataset does not contain any NA values and some columns were removed. To visualize the data, I created a scatterplot to see the distribution of tip amount based on size of the party. Histogram to capture the difference in tip given by male and females as well as another histogram to compare the differences in tip amount based on the different days of the week. Each boxplot has at least 1 outlier, but they are few in numbers and the overall data appears to be consistent.



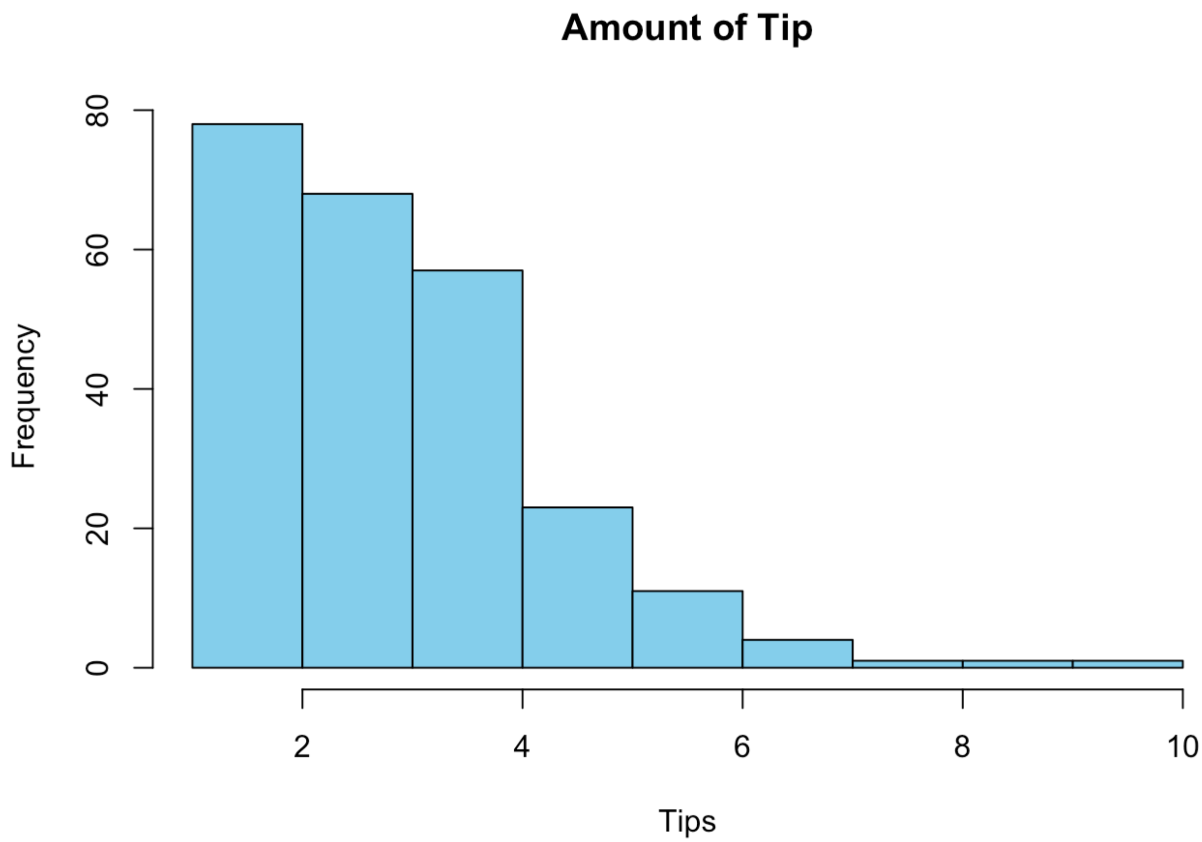
Looking at the scatterplot, we can definitely see that the larger the party size is, the more people tip their waitress/waiters. Party size of 2 to 4 has a similar average tip amount compared to the rest of the party sizes. With 1/individuals tips the least and party size of 6 tips the most. Based on the last histogram, we can see that the distribution is right skewed and not normal.



Based on the boxplot above, it seems that both females and male tips tips the same amount with an average of $3. However, looking at the male section, we can definitely see there's several outliers where they tip very generously compared to females.



The boxplot shows that food servers make an average of $3 per serving on fridays to sundays. However, on thursdays,food servers make less with an average of $2.8.



Looking at the data in more depth, there are definitely several outliers on thursdays and saturdays at which some tip goes up to $10/serving. Therefore, transformation was necessary. Both log and sqrt transformation did transform our y variable into a normal distribution. However, looking at the histograms, log transformation has a better symmetrical distribution.

Statistical Methods

In this study, a Two Sample T-test table is used to compare the means of tip by men and female. Multivariate Linear regression model is used for the second test to find the best fit that would best predict the tip amount. A significance value of 0.05 is used in statistical tests.

The two-sample t-test compares whether the means of two groups are significantly different from each other. This test has the following assumptions: assumes that the observations from each group represent a random sample from the population, assumes that the observations follow a normal distribution, assumes that the observations from the two groups have the same variance. If we have two groups called group 1 and group 2, the two-sided hypothes for the two-sample t-test are:

$$H_0 : \mu_1 = \mu_2$$

$$H_A : \mu_1 \neq \mu_2$$

Before we can test these hypotheses we need to ensure that our assumptions are met. We can test the normality assumption using Shapiro Wilk test. To test the homogeneity (equal) of variance assumption we will use something called Levene's Test. The null hypothesis of the Levene's test is that the variance between the two groups are equal. The alternative hypothesis is that they are not equal. If we fail to reject the null hypothesis of the Levene's test we can pool our variances and use a standard two-sample t-test. (Lab 4).

A multivariate linear model is built to predict the tip amount. The general form of a multivariate linear model is: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$. Y_i is the response variable, and X_i are predictor variables. Because it is hard to explain the whole pattern in the response variable Y_i, combinations of explanatory variables X_i are used to predict the response variable. The assumptions for a general linear model are samples are randomly collected, no collinearity between predictors, some linear relationship between predictors and the response variable, normal and equal variance.

Two Sample T-test table: Comparing the average tip amount between days, gender and group size.

H_0 : The average log tip amount given by male is equal to the average log tip amount given by females.
H_A : The average log tip amount given by male is not equal to the average log tip amount given by females.

I first subsetted the data by gender, male and female. Then I randomly sampled both datasets to produce an equal row number of 80 since the dataset has a larger male input compared to females. Shapiro Wilk test was done to test the normality assumption and both datasets have p-values larger than 0.05. Concluding that both datasets are normally distributed. Next, Levene's test was conducted to test the homogenity of variance and the p-value is also larger than 0.05 meaning that data has equal variance. Since assumption of normality and equal variance are met, I then proceed to do a Two Sample T-test, to which the p-value came out to be 0.1943 > 0.05. Therefore, we failed to reject the null hypothesis and there is no significant difference in the average log tip given by male and female.

Multivariate Linear Model: Predicting tip amount

Three linear regression models were built in order to determine which variables plays a part in predicting the best tip amount. All 3 models, model 1 (log_tip ~ Sex), model 2 (log_tip ~ Sex + day), model 3 (log_tip ~ Sex + day + size) met the normality assumption based on the Q-Qplot. Looking at the Q-Q plot, it seems that not a lot of points deviate from the straight line, the central limit theorem can be applied to conclude that sample data size is large enough so the residuals are normal. All 3 models have no distinct pattern in the residuals at the 0 line in the Residuals vs Fitted plot. Therefore, the assumption of equal variance is met.

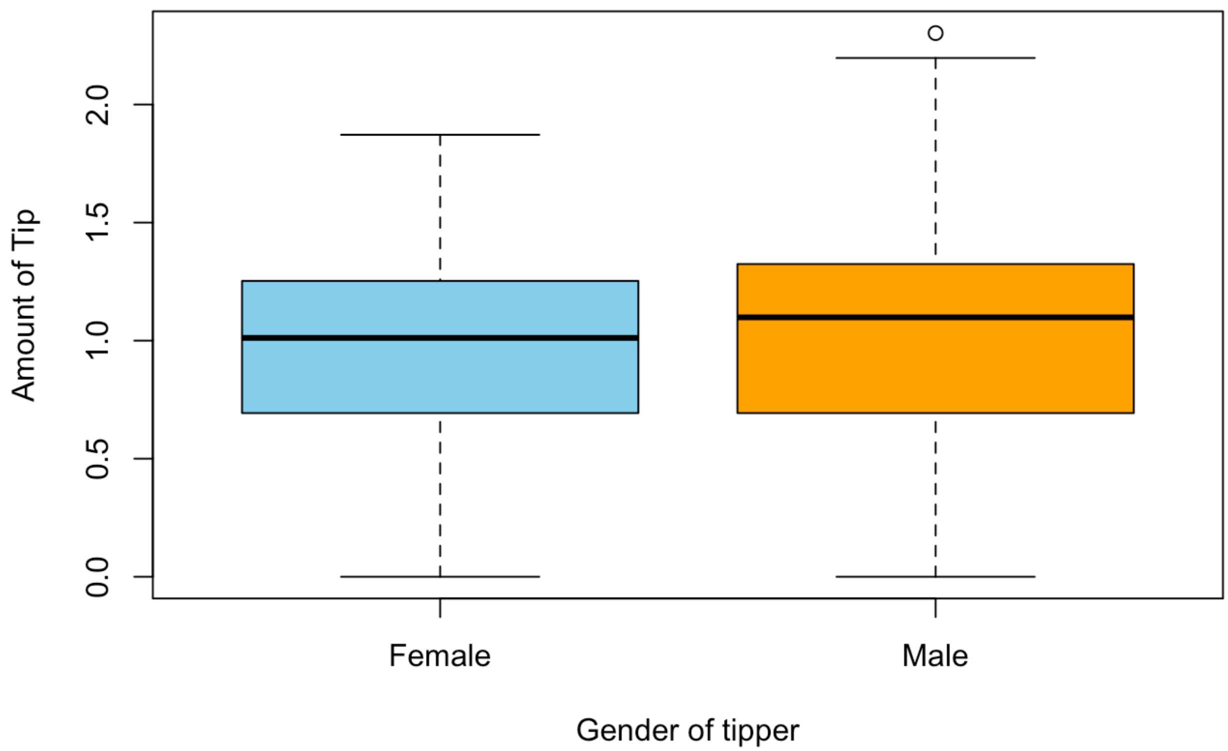
Performance of the Multivariate Linear Model

I examined the effectiveness of each model using R^2 and adjusted R^2. R^2 is the total percentage of variance in the data explained by the model, the value automatically increases as number of parameters increases. Whereas adjusted R^2 is the percentage of variance in the data explained by the model, adjusted for the number of parameters in the model. Once the best model is selected, the model fit is assessed by testing the performance of the trained model. The data set is divided into train set and test set. The test set has fewer than 10% of total data (175 observations) so it is set to be 15 observations. Then, the model is fitted using the train set and this fitted model is used to predict values of the test set. A scatter plot is used to visualize the best model's performance (Lecture 8). Aikake's Information Criterion (AIC), and Bayesian Information Criterion (BIC) also plays a huge role in model selection. BIC places a higher penalty on the number if predictors than AIC. Both AIC and BIC indicates a better model when their values are smaller. Therefore, the best model is selected by choosing a model with the lowest AIC and BIC value with the highest R^2 and adjusted R^2 values.

Once the best model has been selected, the model fit is assessed by testing the performance of the trained data on the test data. The data is divided into train set and test set. The test set has fewer than 10% of total data so it is set to be 15 observations. Then, the model is fitted using the train set and this fitted model is used to predict values of the test set. A scatter plot is used to visualize the best model's performance.

Results

Two Sample T-Test




```
##
##  Two Sample t-test
##
## data:  male$log_tip and female$log_tip
## t = 1.5621, df = 158, p-value = 0.1203
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02813286  0.24094016
## sample estimates:
## mean of x mean of y
## 1.0599402 0.9535365
```

The box-plot above visualizes the difference in the log tip amount given by male and female. The two sample t-test has a p-value of 0.1943 which is larger than 0.05. Therefore, we failed to reject the null hypothesis and there is no significant difference in the average log tip given by male and female.

Best Multivariate Linear model: fit3_tip

Looking at the table below, fit 3 has the lowest AIC, BIC values with the highest R^2 and adjusted R^2 values. Therefore, fit 3 is the best model for this study.

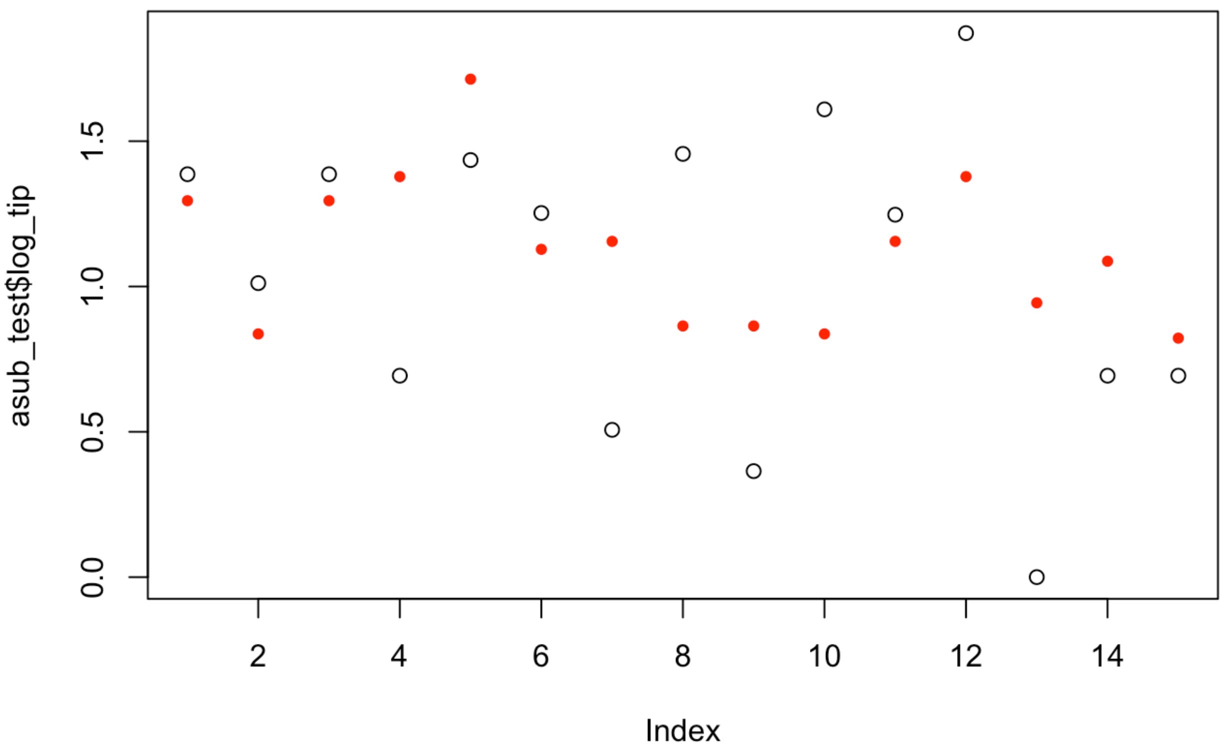
	df	AIC	BIC	rsq	adj_rsq
fit1_tip	3	290.97	301.46	0.01	0.00
fit2_tip	6	291.24	312.22	0.03	0.01
fit3_tip	7	234.00	258.48	0.24	0.22

```
##
## Call:
## lm(formula = log_tip ~ sex + day + size, data = tips)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.89671 -0.23652  0.00908  0.25076  1.21291
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.465943   0.107484   4.335 2.15e-05 ***
## sexMale      0.023831   0.052910   0.450    0.653
## daySat      -0.046257   0.098283  -0.471    0.638
## daySun       0.009126   0.101231   0.090    0.928
## dayThur     -0.073025   0.101294  -0.721    0.472
## size         0.215385   0.026632   8.087 3.11e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3846 on 238 degrees of freedom
## Multiple R-squared:  0.2386, Adjusted R-squared:  0.2226
## F-statistic: 14.92 on 5 and 238 DF,  p-value: 9.765e-13
```

From the summary, we can see that size of the party has p-values that are smaller than 0.05. Meaning that said variable plays a significant role in predicting tip amount. The multivariate linear model can be written in the following general form $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon_i$ or $\text{log\_tip} = 0.4659 + 0.0238(\text{sexMale}) - 0.0463(\text{daySat}) + 0.0091(\text{daySun}) - 0.0730(\text{dayThur}) + 0.2154(\text{size}) + \epsilon_i$.

Prediction of Best model: fit3_tip

229 observations were used to train the best model and 15 observations were used to test out the fitted model. The observation vs prediction result is shown using a scatter plot below. The red dots are the predicted values and the non colored circles are the actual values (log_tip). Looking at the scatter plot, we can see that some predicted values are accurate as it overlaps the actual values. Therefore, we can conclude that fit 3 is the best model. Meaning that size of the party, male, saturday, sunday and thursday are the variables used to determine or predict the tip amount.



Discussion

Real Life applications

Realistically speaking, there is no definitive way in predicting the amount of tip that a person would give. It could depend on the mood of the person that day. Is he/she feeling generous? are they strapped in cash? Or are they just following the 15% 18% tipping rule? We will never know.

Future Work

Future research in regard to patron's age or age range could greatly expand the collective understanding of tipping. While age range is an exceptionally subjective variable, it is in the authors opinion as an experienced server that someone's tentative age can be a very telling characteristic for how you, as a server, will be treated as well as tipped. Understanding the limitations of the variable is to be noted; however it should not be overlooked (Nelson 2017). On top of that, data that was used on this study may be biased. Data was collected by a single food server at a certain restaurant. Meaning this data does not represent a majority of the tips received by food servers. In other words, the data collected was not random and could be biased.

References

Azar O. H. (2005). "Who do we tip and why? An empirical investigation", Applied Economics, Vol. 37, No. 16, pp. 1871-1879.

Azar O. H. (2007). "The social norm of tipping: A review", Journal of Applied Social Psychology, Vol. 37, No. 2, pp. 380-402.

Conlin T. (2003). "The norm of restaurant tipping", Journal of Economic Behavior & Organization, Vol. 52, No. 3, pp. 297-321.

Lynn M. (1997). “Tipping customs and status seeking: A crosscountry study”, International Journal of Hospitality Management, Vol. 16, pp. 221-224.
Lynn M., Zinkhan G. M. and Harris J. (1993). “Consumer tipping: A crosscountry study”, Journal of Consumer Research, Vol. 20, pp. 478-488.
Nelson, M. (2017). A case study in tipping: An economic anomaly. Crossing Borders: A Multidisciplinary Journal of Undergraduate Scholarship, 2(1). https://doi.org/10.4148/2373-0978.1021 (https://doi.org/10.4148/2373-0978.1021)
Popa, Monica, and Shayne Hurd. (2016). “The impact of sexual orientation on consumers’ tipping behavior.” Journal of Business and Economics 7.1: 94-104.

Appendix

```
# library
library(kableExtra)
library(ggplot2)
library(car)
library(dplyr)
library(knitr)
library(psych)
library(tidyverse)

# sets worki# sets working directory
setwd('~/Desktop/EEMB 146')
# load data
tips <- read.csv('tips.csv')

# checks dimension of data
dim(tips) # 244 rows and 7 columns
```

```
## [1] 244   7
```

```
str(tips) # 7 variables
```

```
## 'data.frame':    244 obs. of  7 variables:
##  $ total_bill: num  17 10.3 21 23.7 24.6 ...
##  $ tip       : num  1.01 1.66 3.5 3.31 3.61 4.71 2 3.12 1.96 3.23 ...
##  $ sex       : chr  "Female" "Male" "Male" "Male" ...
##  $ smoker    : chr  "No" "No" "No" "No" ...
##  $ day       : chr  "Sun" "Sun" "Sun" "Sun" ...
##  $ time      : chr  "Dinner" "Dinner" "Dinner" "Dinner" ...
##  $ size      : int  2 3 3 2 4 4 2 4 2 2 ...
```

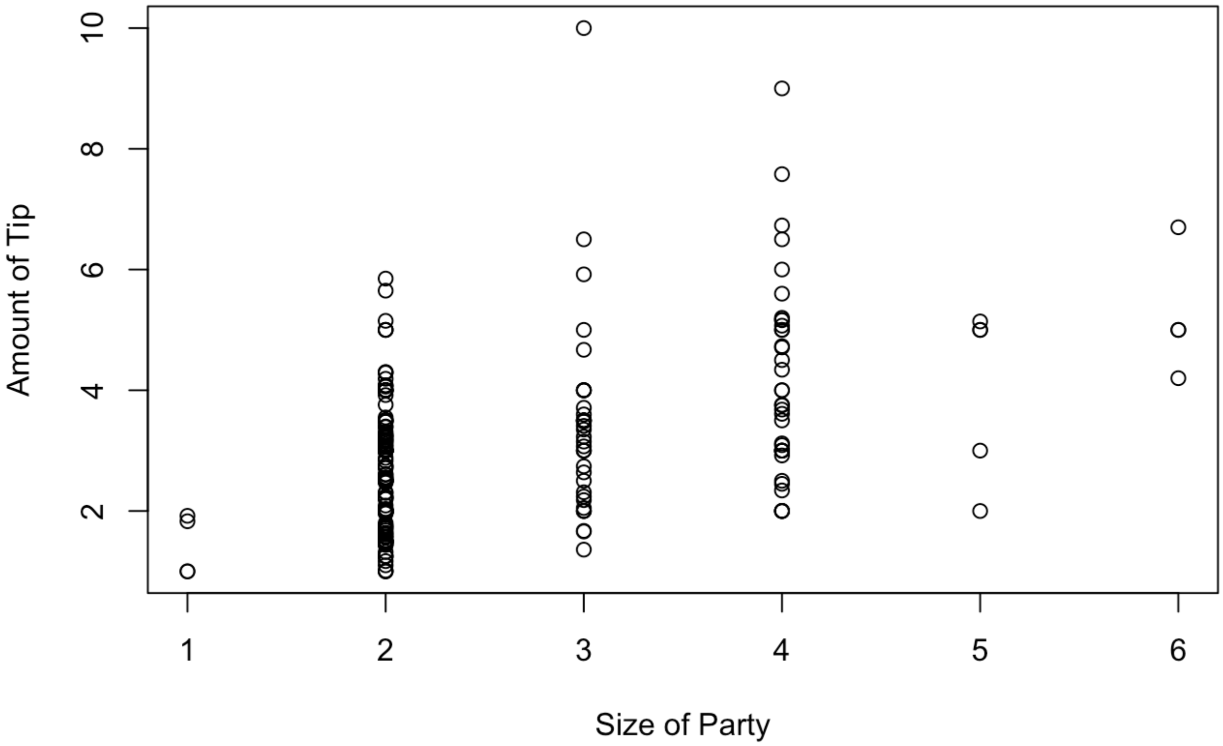
```
summary(tips)
```

```
##    total_bill         tip            sex               smoker         
##  Min.   : 3.07   Min.   : 1.000   Length:244         Length:244        
##  1st Qu.:13.35   1st Qu.: 2.000   Class :character   Class :character  
##  Median :17.80   Median : 2.900   Mode  :character   Mode  :character  
##  Mean   :19.79   Mean   : 2.998                                        
##  3rd Qu.:24.13   3rd Qu.: 3.562                                        
##  Max.   :50.81   Max.   :10.000                                        
##      day                time                size     
##  Length:244         Length:244         Min.   :1.00  
##  Class :character   Class :character   1st Qu.:2.00  
##  Mode  :character   Mode  :character   Median :2.00  
##                                        Mean   :2.57  
##                                        3rd Qu.:3.00  
##                                        Max.   :6.00
```

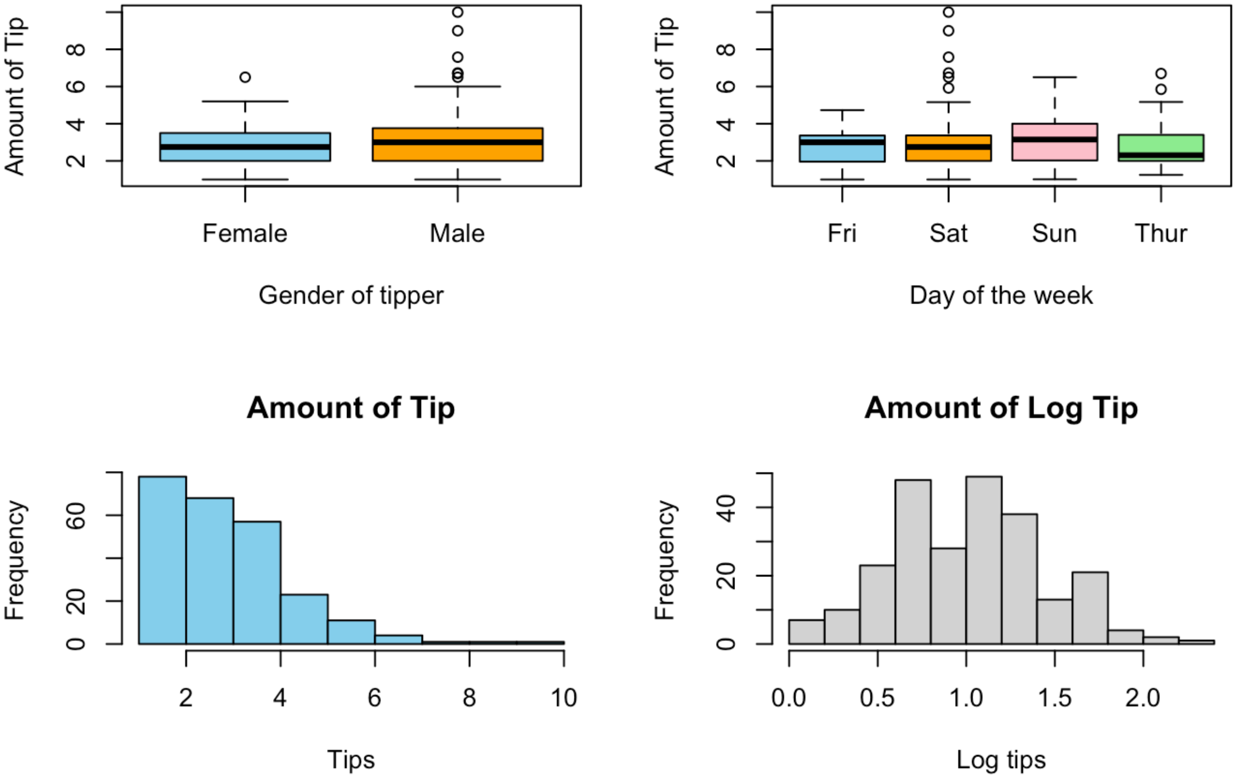
```
# check if there's any NA values
sum(is.na(tips)) # 0
```

```
## [1] 0
```

```
# using the select function to pull specific columns into a NEW dataframe
tips <- tips %>%
    select(tip, sex, day, size)

# scatterplot
plot(tip ~ size, data = tips, xlab='Size of Party', ylab='Amount of Tip')
```



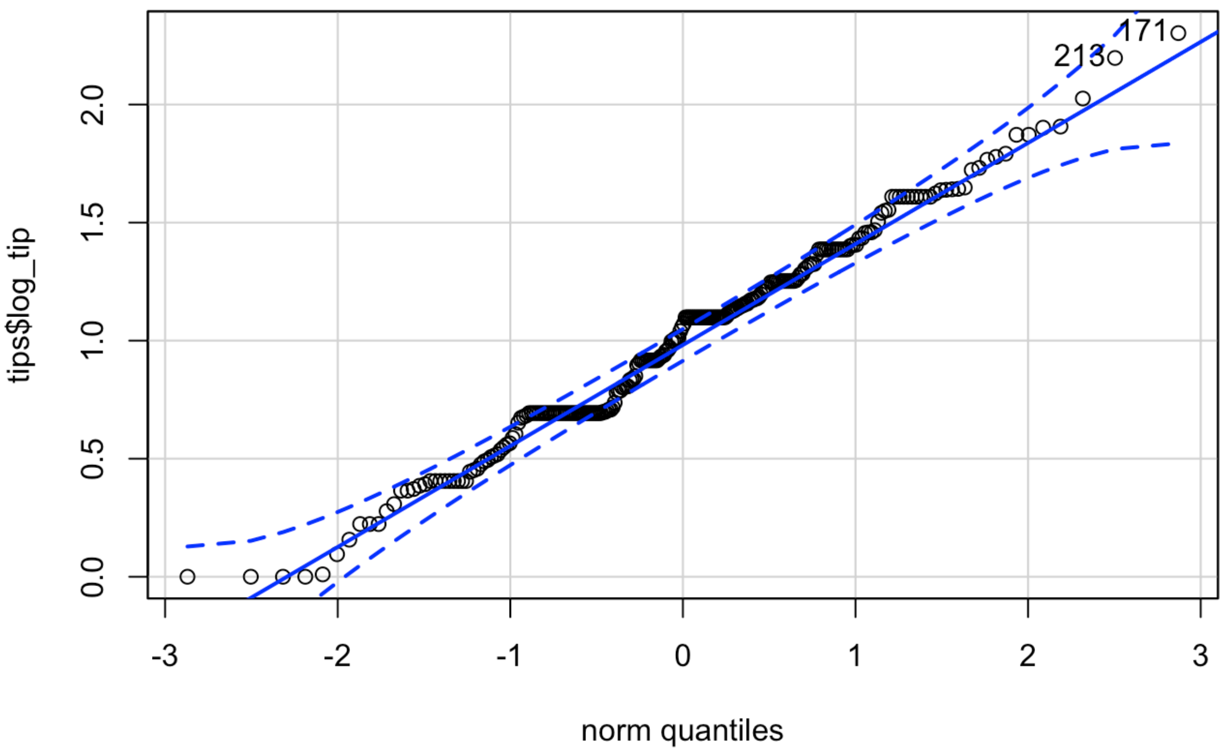

```
par(mfrow=c(2,2))

# boxplot
boxplot(tip ~ sex, data = tips, xlab='Gender of tipper', ylab='Amount of Tip',
        col = c('Sky Blue', 'Orange'))

# boxplot
boxplot(tip ~ day, data = tips, xlab='Day of the week', ylab='Amount of Tip',
        col = c('Sky Blue', 'Orange', 'Pink', 'Light Green'))

# histogram to check normality of our Y variable
hist(tips$tip, xlab='Tips', main='Amount of Tip', col=c('Sky Blue'))

# Y variable is right skewed and not symmetrical.
# Meaning it's not normally distributed.

# log transformation
tips$log_tip <- log(tips$tip)
hist(tips$log_tip, xlab='Log tips', main='Amount of Log Tip')
```




```
# data looks normal after log transformation

par(mfrow=c(1,1))

# Shapiro-Wilks test
shapiro.test(tips$log_tip) # p value = 0.05621 > 0.05, failed to reject null hypothesis
```

```
##
##  Shapiro-Wilk normality test
##
## data:  tips$log_tip
## W = 0.98885, p-value = 0.05621
```

```
# data is normal
# qqPlot
qqPlot(tips$log_tip) # Y variable looks normally distributed
```




```
## [1] 171 213
```

```
# FIRST TEST
# Two Sample T-test to compare means
# include log transformed tip
tips <- tips %>%
    select(sex, day, size, log_tip)

# subset data to males and females
male <- subset(tips, tips$sex == "Male")
female <- subset(tips, tips$sex == "Female")

# randomly sample data
male <- sample_n(male, 80)
female <- sample_n(female, 80)

# shapiro test
shapiro.test(male$log_tip) # > 0.05
```

```
##
##  Shapiro-Wilk normality test
##
## data:  male$log_tip
## W = 0.97186, p-value = 0.07454
```

```r
shapiro.test(female$log_tip) # > 0.05
```

```
##
##  Shapiro-Wilk normality test
##
## data:  female$log_tip
## W = 0.97624, p-value = 0.1419
```

```r
# levene test to check for equal variance
leveneTest(tips$log_tip, tips$sex) # p > 0.05
```

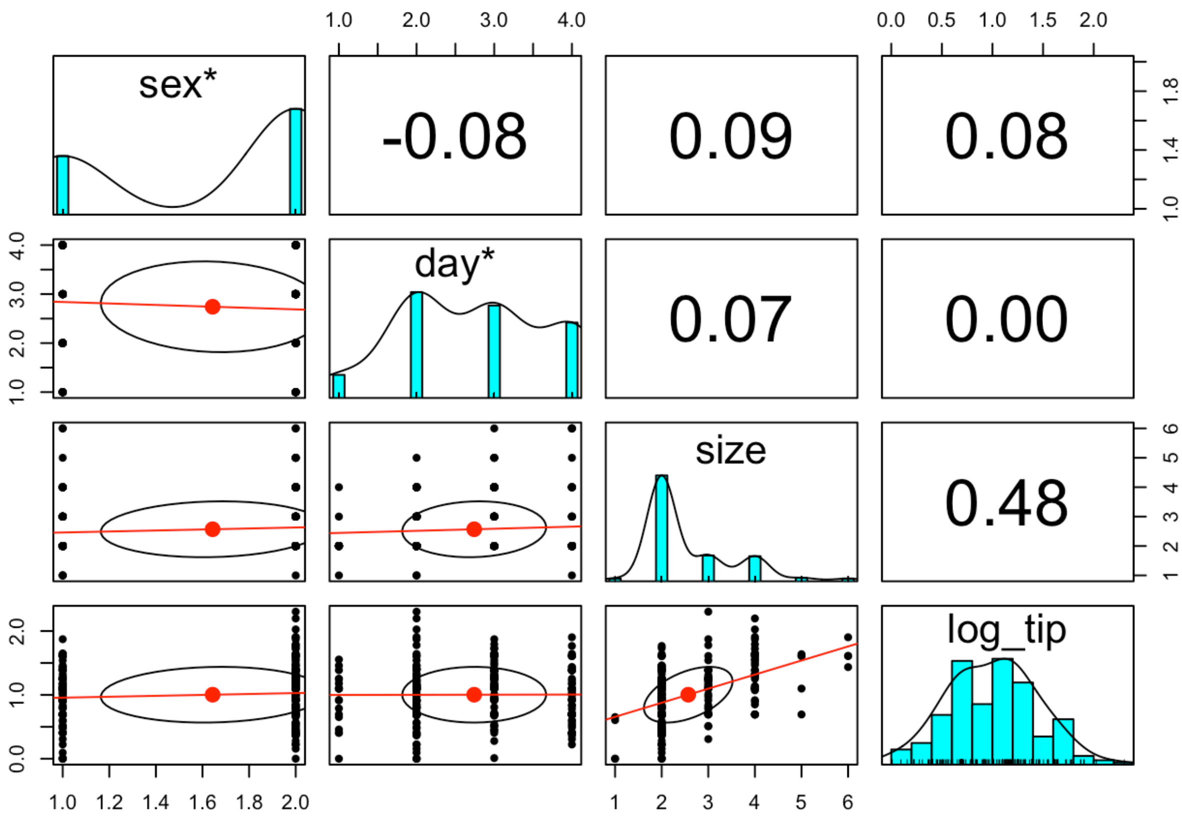
	Df <int>	F value <dbl>	Pr(>F) <dbl>
group	1	0.2756858	0.6000236
	242	NA	NA

2 rows

```r
# two sample t-test
t.test(male$log_tip, female$log_tip, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  male$log_tip and female$log_tip
## t = 0.9384, df = 158, p-value = 0.3495
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.0723243  0.2032581
## sample estimates:
## mean of x mean of y
## 1.0446142 0.9791473
```

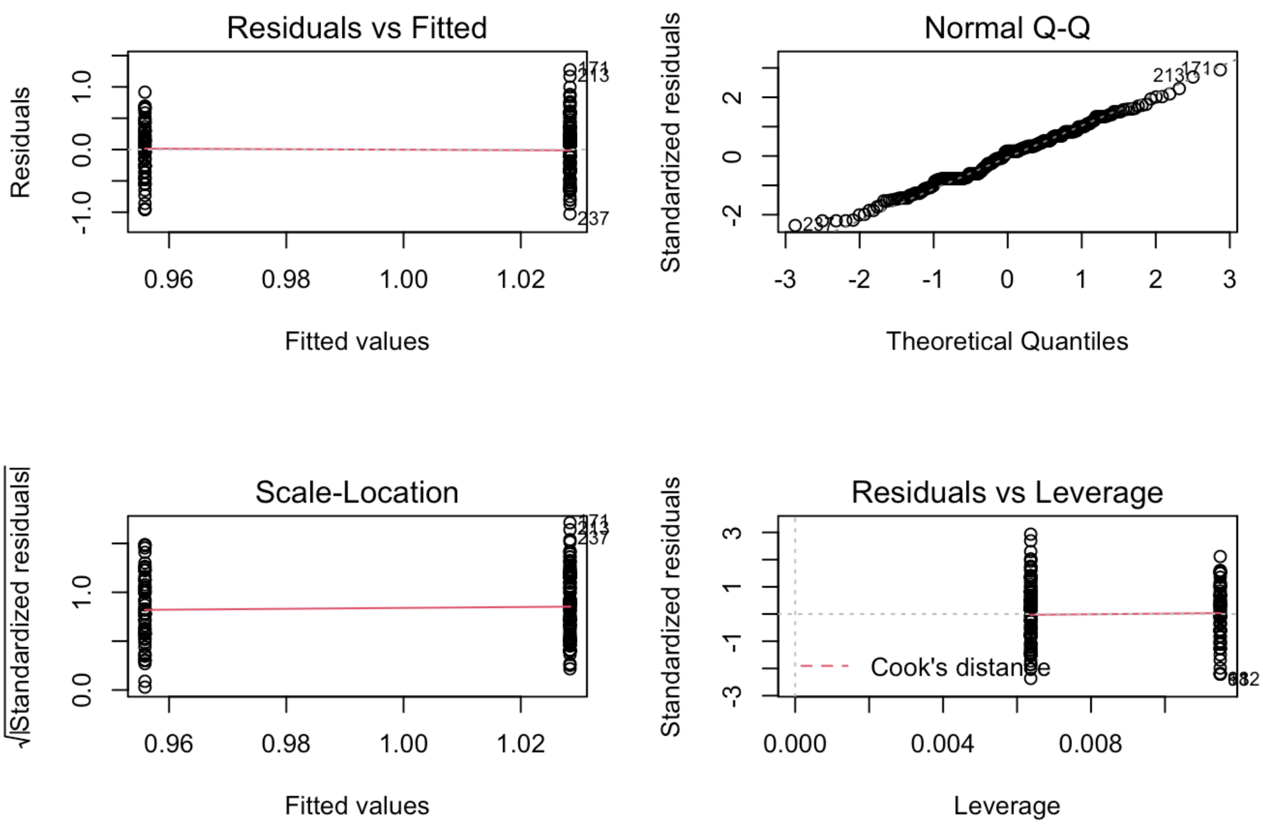
```r
# analyzing the two sample t-test
# p-value = 0.1943 > 0.05
# failed to reject null hypothesis that average tip given by male is equal tip average tip give
n by females

# SECOND TEST
# check for collinearity
pairs.panels(tips, lm = TRUE, cor = T)
```




```
# build linear regression models
fit1_tip <- lm(log_tip~sex, data = tips)
fit2_tip <- lm(log_tip~sex + day, data = tips) # best model
fit3_tip <- lm(log_tip~sex + day + size, data = tips)

par(mfrow=c(2,2))

summary(fit1_tip)
```

```
##
## Call:
## lm(formula = log_tip ~ sex, data = tips)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.02839 -0.33525  0.07022  0.29624  1.27419
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.95588    0.04671  20.465   <2e-16 ***
## sexMale      0.07251    0.05823   1.245    0.214
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4357 on 242 degrees of freedom
## Multiple R-squared:  0.006367,    Adjusted R-squared:  0.002261
## F-statistic: 1.551 on 1 and 242 DF,  p-value: 0.2142
```

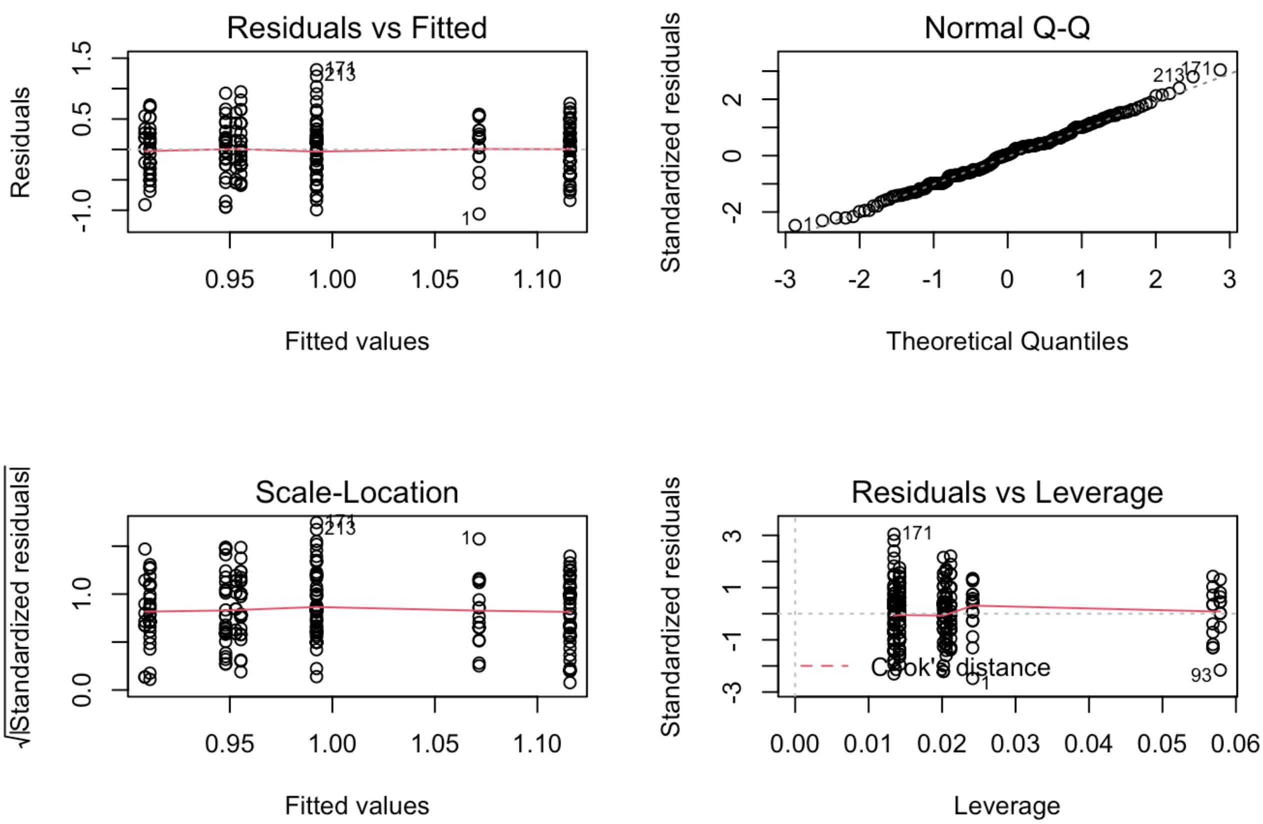
```
plot(fit1_tip)
```




```
summary(fit2_tip)
```

```
##
## Call:
## lm(formula = log_tip ~ sex + day, data = tips)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06162 -0.29914  0.01597  0.26882  1.31030
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.908594   0.104225   8.718 4.86e-16 ***
## sexMale     0.044332   0.059546   0.745    0.457
## daySat      0.039364   0.110092   0.358    0.721
## daySun      0.162975   0.112025   1.455    0.147
## dayThur     0.002443   0.113644   0.021    0.983
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4333 on 239 degrees of freedom
## Multiple R-squared:  0.02941,    Adjusted R-squared:  0.01316
## F-statistic:  1.81 on 4 and 239 DF,  p-value: 0.1275
```

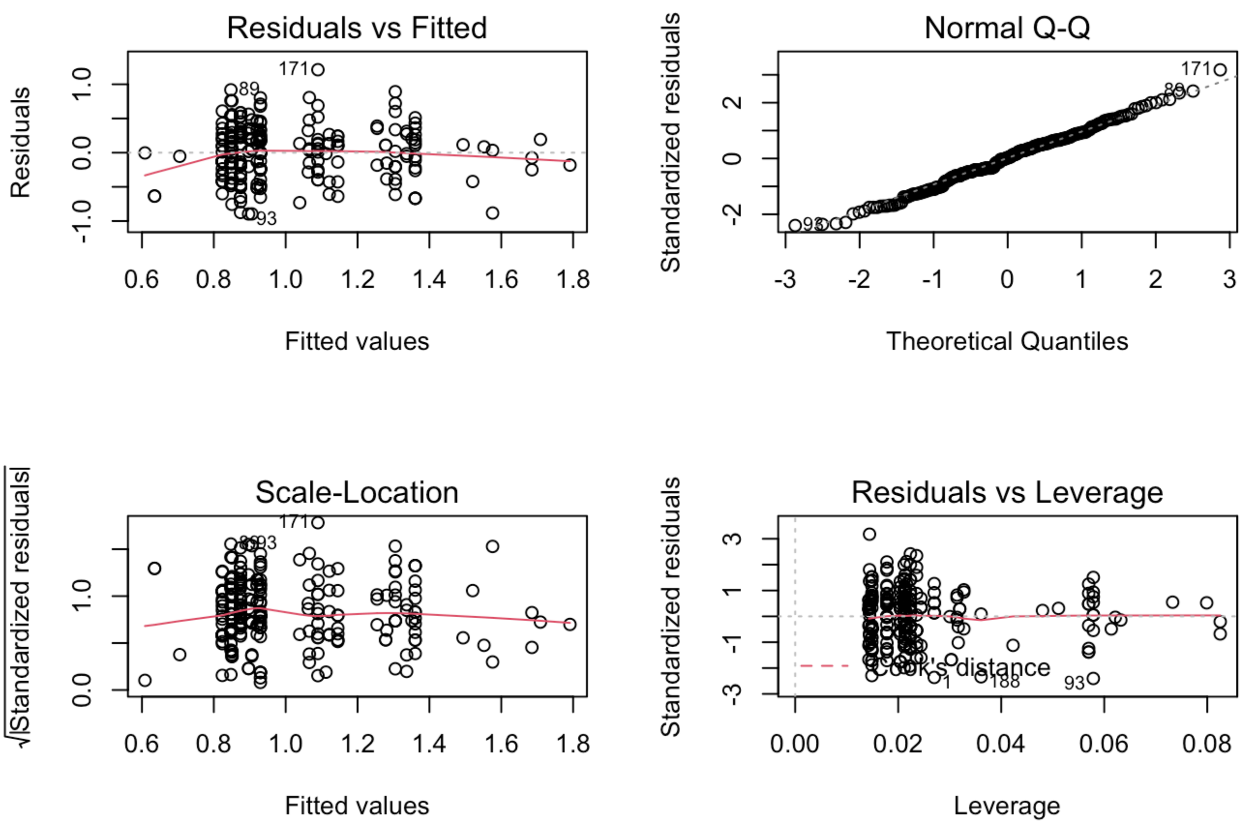
```
plot(fit2_tip)
```




```
summary(fit3_tip)
```

```
##
## Call:
## lm(formula = log_tip ~ sex + day + size, data = tips)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.89671 -0.23652  0.00908  0.25076  1.21291
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.465943   0.107484   4.335 2.15e-05 ***
## sexMale      0.023831   0.052910   0.450    0.653
## daySat      -0.046257   0.098283  -0.471    0.638
## daySun       0.009126   0.101231   0.090    0.928
## dayThur     -0.073025   0.101294  -0.721    0.472
## size         0.215385   0.026632   8.087 3.11e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3846 on 238 degrees of freedom
## Multiple R-squared:  0.2386, Adjusted R-squared:  0.2226
## F-statistic: 14.92 on 5 and 238 DF,  p-value: 9.765e-13
```

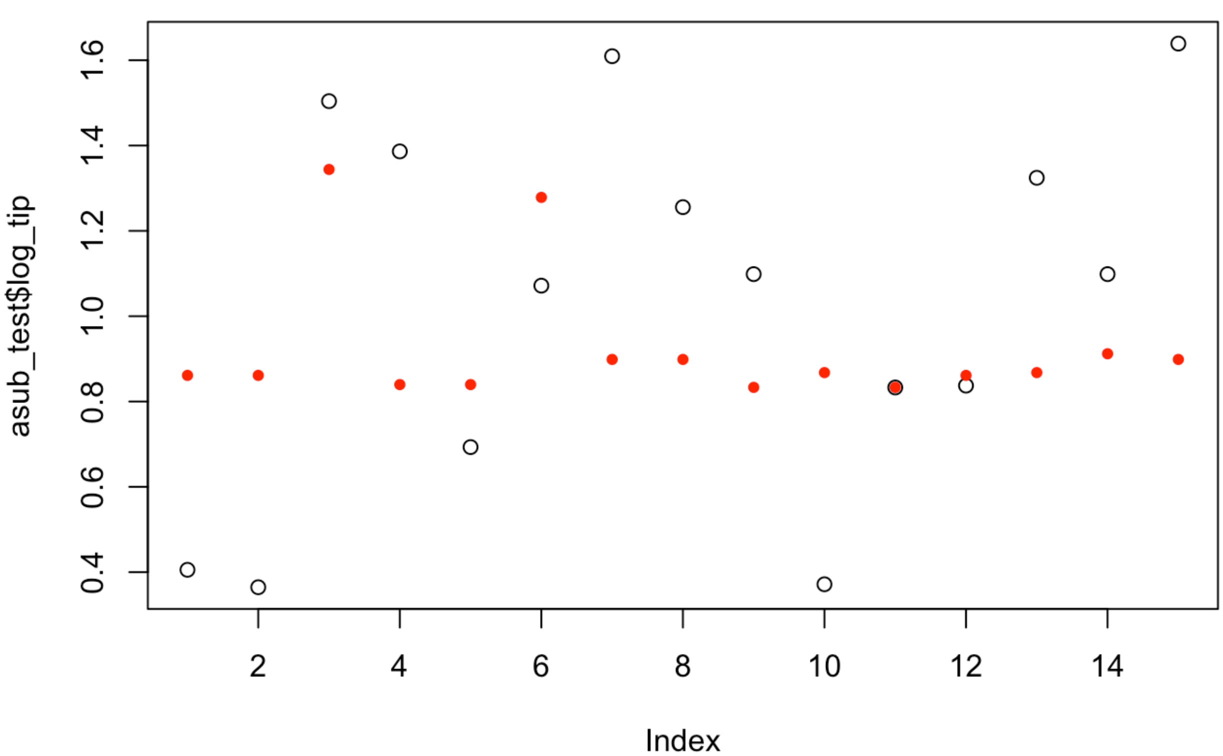
```
plot(fit3_tip)
```




```
# has the best multiple R^2 value of 0.2386
# 23% of the variation in tip amount can be explained by sex
# of tipper, the day of the week and size of party.

# calculate AIC of each model
result <- AIC(fit1_tip, fit2_tip, fit3_tip)

# add other metrics to table
models <- list(fit1_tip, fit2_tip, fit3_tip)
result$BIC <- sapply(models, BIC) # add a column for BIC to the results
model_summary <- lapply(models, summary)

for(i in 1:length(models)){
  result$rsq[i] <- model_summary[[i]]$r.squared
  result$adj_rsq[i] <- model_summary[[i]]$adj.r.squared
}
kable(result, digits = 2, align = "c")
```

	df	AIC	BIC	rsq	adj_rsq
fit1_tip	3	290.97	301.46	0.01	0.00
fit2_tip	6	291.24	312.22	0.03	0.01
fit3_tip	7	234.00	258.48	0.24	0.22

```r
# fit2_tip is the best model
# with the lowest AIC and BIC value
# and highest adjusted R^2 value

par(mfrow=c(1,1))

# separate your data into a training set (most of the data) and a test set (a few observations,
or <10% of rows)
splitter <- sample(1:nrow(tips), 15, replace = F) # pick 15 random rows from tips to reserve as
test data
asub_train <- tips[-splitter,] # leave those rows out of the training data
asub_test <- tips[splitter,] # use them to create a set of test data

fit_3var_split <- lm(log_tip~sex + day + size, data = asub_train)
prediction <- predict(fit_3var_split, asub_test)

plot(asub_test$log_tip, pch=1) # plot the actual test data values
points(prediction, pch=20, col = "red") # plot the model predictions for those points
```




```r
# cite packages
citation('kableExtra')
```

```
##
## To cite package 'kableExtra' in publications use:
##
##   Hao Zhu (2021). kableExtra: Construct Complex Table with 'kable' and
##   Pipe Syntax. R package version 1.3.4.
##   https://CRAN.R-project.org/package=kableExtra
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {kableExtra: Construct Complex Table with 'kable' and Pipe Syntax},
##     author = {Hao Zhu},
##     year = {2021},
##     note = {R package version 1.3.4},
##     url = {https://CRAN.R-project.org/package=kableExtra},
##   }
```

```
citation('car')
```

```
##
## To cite the car package in publications use:
##
##   John Fox and Sanford Weisberg (2019). An {R} Companion to Applied
##   Regression, Third Edition. Thousand Oaks CA: Sage. URL:
##   https://socialsciences.mcmaster.ca/jfox/Books/Companion/
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     title = {An {R} Companion to Applied Regression},
##     edition = {Third},
##     author = {John Fox and Sanford Weisberg},
##     year = {2019},
##     publisher = {Sage},
##     address = {Thousand Oaks {CA}},
##     url = {https://socialsciences.mcmaster.ca/jfox/Books/Companion/},
##   }
```

```
citation('knitr')
```

```
##
## To cite the 'knitr' package in publications use:
##
##   Yihui Xie (2021). knitr: A General-Purpose Package for Dynamic Report
##   Generation in R. R package version 1.31.
##
##   Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition.
##   Chapman and Hall/CRC. ISBN 978-1498716963
##
##   Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible
##   Research in R. In Victoria Stodden, Friedrich Leisch and Roger D.
##   Peng, editors, Implementing Reproducible Computational Research.
##   Chapman and Hall/CRC. ISBN 978-1466561595
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.
```

```
citation('psych')
```

```
##
## To cite the psych package in publications use:
##
##   Revelle, W. (2022) psych: Procedures for Personality and
##   Psychological Research, Northwestern University, Evanston, Illinois,
##   USA, https://CRAN.R-project.org/package=psych Version = 2.2.5.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {psych: Procedures for Psychological, Psychometric, and Personality Research},
##     author = {William Revelle},
##     organization = { Northwestern University},
##     address = { Evanston, Illinois},
##     year = {2022},
##     note = {R package version 2.2.5},
##     url = {https://CRAN.R-project.org/package=psych},
##   }
```

```
citation('tidyverse')
```

```
##
##   Wickham et al., (2019). Welcome to the tidyverse. Journal of Open
##   Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {Welcome to the {tidyverse}},
##     author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy
D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henry an
d Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bache and
Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spinu and Ko
hske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani},
##     year = {2019},
##     journal = {Journal of Open Source Software},
##     volume = {4},
##     number = {43},
##     pages = {1686},
##     doi = {10.21105/joss.01686},
##   }
```

```
citation('ggplot2')
```

```
##
## To cite ggplot2 in publications, please use:
##
##   H. Wickham. ggplot2: Elegant Graphics for Data Analysis.
##   Springer-Verlag New York, 2016.
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     author = {Hadley Wickham},
##     title = {ggplot2: Elegant Graphics for Data Analysis},
##     publisher = {Springer-Verlag New York},
##     year = {2016},
##     isbn = {978-3-319-24277-4},
##     url = {https://ggplot2.tidyverse.org},
##   }
```

```
citation('dplyr')
```

```
##
## To cite package 'dplyr' in publications use:
##
##   Hadley Wickham, Romain François, Lionel Henry and Kirill Müller
##   (2021). dplyr: A Grammar of Data Manipulation. R package version
##   1.0.7. https://CRAN.R-project.org/package=dplyr
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {dplyr: A Grammar of Data Manipulation},
##     author = {Hadley Wickham and Romain François and Lionel Henry and Kirill Müller},
##     year = {2021},
##     note = {R package version 1.0.7},
##     url = {https://CRAN.R-project.org/package=dplyr},
##   }
```