Predicting Yelp Ratings Using Multilevel Ordinal Regression

Albert Ding

December 2018

Abstract

Yelp is the leading provider of online business reviews crowdsourced as user generated content. When consumers are contemplating purchasing decisions, Yelp can heavily influence the buyer's final decision. The impact of these ratings, scored from one to five stars, is especially pronounced within the restaurant industry, where 40% of the company's 170 million reviews are focused.

Given this phenomenon, I leveraged the restaurant ratings of Yelp to examine what factors influence the rating from an individual review. I explored several hypotheses and estimated the relationship between ratings and several independent variables using linear regression, ordinal multinomial regression, and multilevel linear and ordinal multinomial regression.

The onus of conducting a multilevel model arose from my discovery that ratings of the same business are significantly clustered. Based on the model estimates, I presented evidence that restaurants in the second price category, or two dollar signs on Yelp, have lower ratings, and reviews written by on average more satisfied users tend to be more positive, the strongest observed relationship of all my analysis. The results suggest that restaurant management and patrons alike should pay attention to features other than descriptions and pictures of food on Yelp and should consider who is writing the review before giving it weight. Given the limited findings, a plethora of topics remain to be explored linking the relationship between restaurant rating and information on reviewer demographic and other restaurant features.

Introduction

Since its founding in 2004, Yelp has aimed to provide a digital form for word-of-mouth recommendations. Replicating the positive / negative sentiment of business reviews, Yelp requires a rating from one to five in one star increments for each review. These ratings are then averaged in aggregate for restaurants and shown in half star increments.

Given the popularity and mindshare of Yelp among consumers and restaurants, these ratings leave a strong impression. Yelp boasts over 173 million monthly active users as of Q3 2018, and among consumers, Yelp, Facebook, and Google are the most trusted review sites according to Uberall, a location marketing software company. (1)

Restaurants are keenly aware of these facts; among small and medium size businesses, Yelp ranks as the fourth most popular channel for online advertising with 22% paying in one form or another. A frequently-cited paper written by Professor Michael Luca of Harvard Business School even proposes a causal inference on sales in proclaiming, "a one-star increase in Yelp rating leads to a 5-9 percent increase in revenue." The implications of Yelp ratings on both consumer perception and business results cannot be understated. (2) (3)

In exploring the nuances of reviewer ratings, this project sheds light on how reviewer satisfaction differs. These learnings may prove helpful to anyone reading a restaurant description or review and wondering how to interpret them.

Method

Using data from the twelfth annual Yelp Academic Challenge, I explored several hypotheses about restaurant experiences. I chose this dataset because Yelp is the largest source of restaurant reviews worldwide, and their publicly available data are likely the most representative of American restaurant reviews available, especially considering 90% of Yelp's user base lives in the US.

I first explored several potential factors contributing to positive restaurant ratings through univariate and correlational analyses. Next, I predicted reviewer satisfaction as a function of several explanatory variables using linear regression. I noticed that my results were mismatched because I was modeling an ordinal response as a continuous outcome. Nevertheless, I referenced the linear model throughout this paper and preserved its results in the appendix.

I then predicted ratings as a five-category ordinal response using ordinal multinomial regression, which agreed directionally with the linear models. Finally, I show that there is substantial clustering of rating scores within reviews of the same restaurant, indicating that multilevel models are more appropriate than single level ones. Based on this finding, I predicted satisfaction as a multilevel ordinal regression allowing varying intercepts within groupings for the same business. In the appendix, I have also followed through with a multilevel linear model for comparative purposes.

Data

To do this analysis, I first parsed the JSON files publicly available at www.yelp.com/dataset using base R and the jsonlite packages. After including only restaurants, removing missing values, and excluding businesses with less than 10 reviews, I narrowed the dataset down to 180,000 reviews or so. From there, I took a random subset of 10,000 reviews to improve computation speed.

I created independent variables for the years the reviewer has been writing on Yelp, his or her average rating conferred in the past, and the length of the review. On the side of restaurants, I generated independent variables from the price category of the restaurant, whether the restaurant has TVs, and whether the restaurant is open 24 hours a day. Finally, I converted variables to their appropriate types, such as converting the ratings to categorical for the ordinal logistic regression and centering and scaling the continuous predictors by subtracting their mean and dividing by one standard deviation. Although the interpretability of the continuous predictors were diminished along the way, their effect size became more directly comparable.

I decided to put my analysis of the relationship between review word count and rating into the appendix. I found that there was a generally negative relationship, as longer reviews tend to be less positive, but I decided not to include this feature in the regression analysis. My reasoning is that word count is more of a consequence of restaurant experience than anything about the restaurant or the reviewers themselves.

Results

Distribution of Review Stars

Looking at the distribution of review ratings, you can see that the majority of the 189,000 total restaurant reviews in the dataset prior to sampling are 4's and 5's.
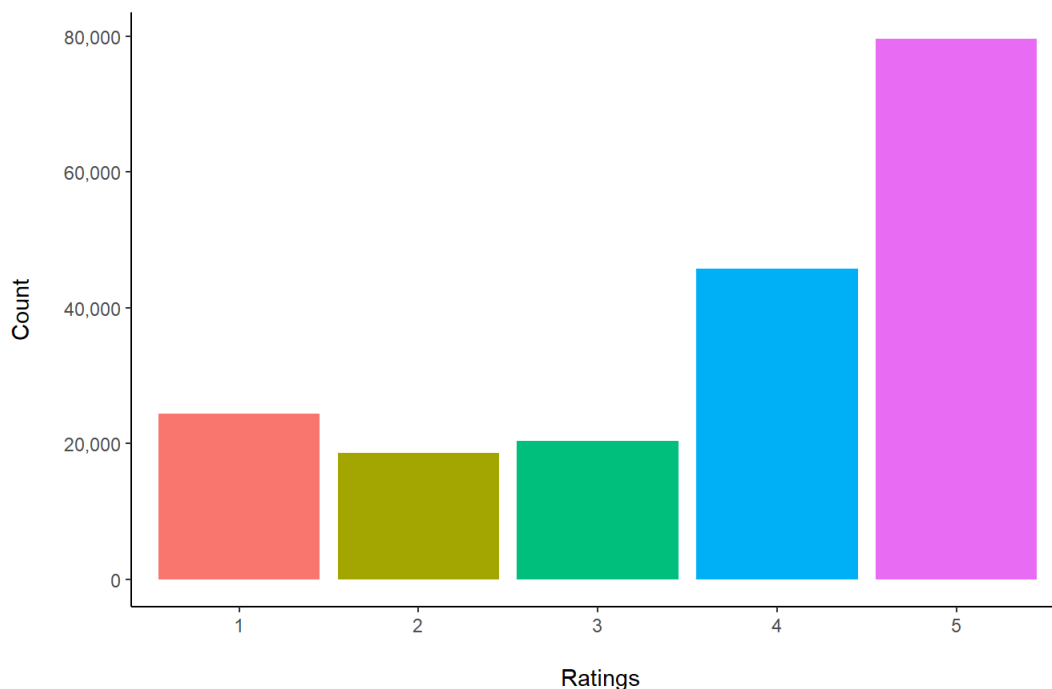
User's Average Review Stars

Reviews are strongly impacted by the author, as reviews by generally positive users are much more likely to be positive. Note that the size of the points below are proportional to sample size observed for each.

Restaurant Ratings of Users by Mean User Rating Bins
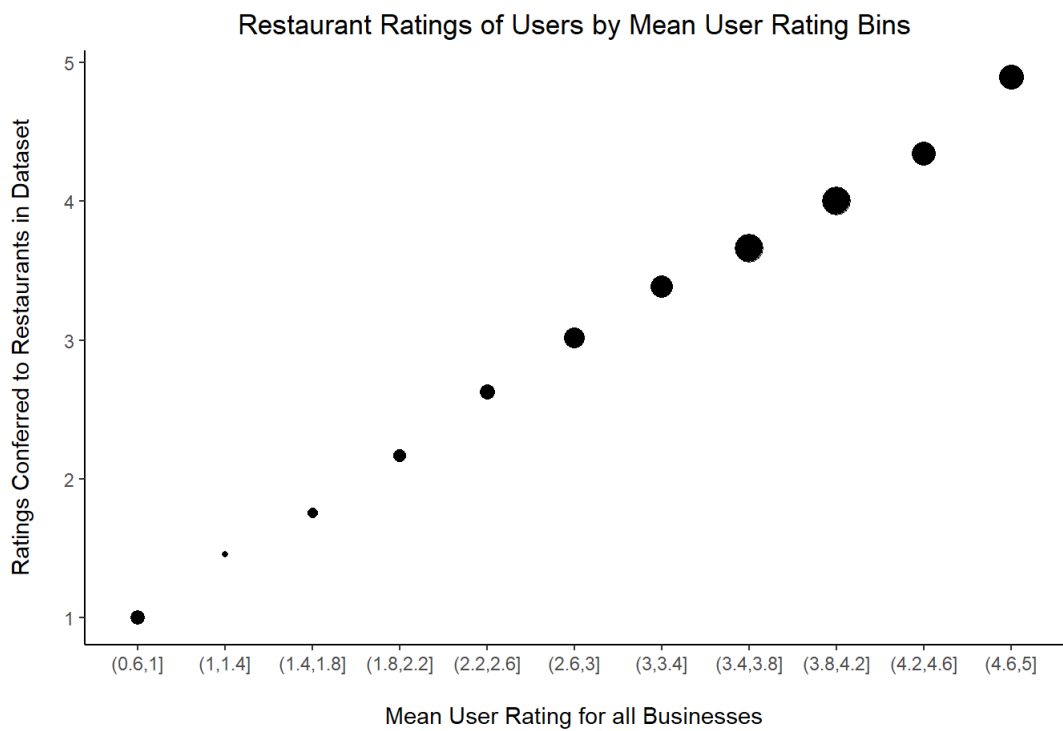
Years Reviewer on Yelp Versus Ratings

There appears to be a weak relationship between years Yelping and ratings.

Ratings by Users' Years on Yelp
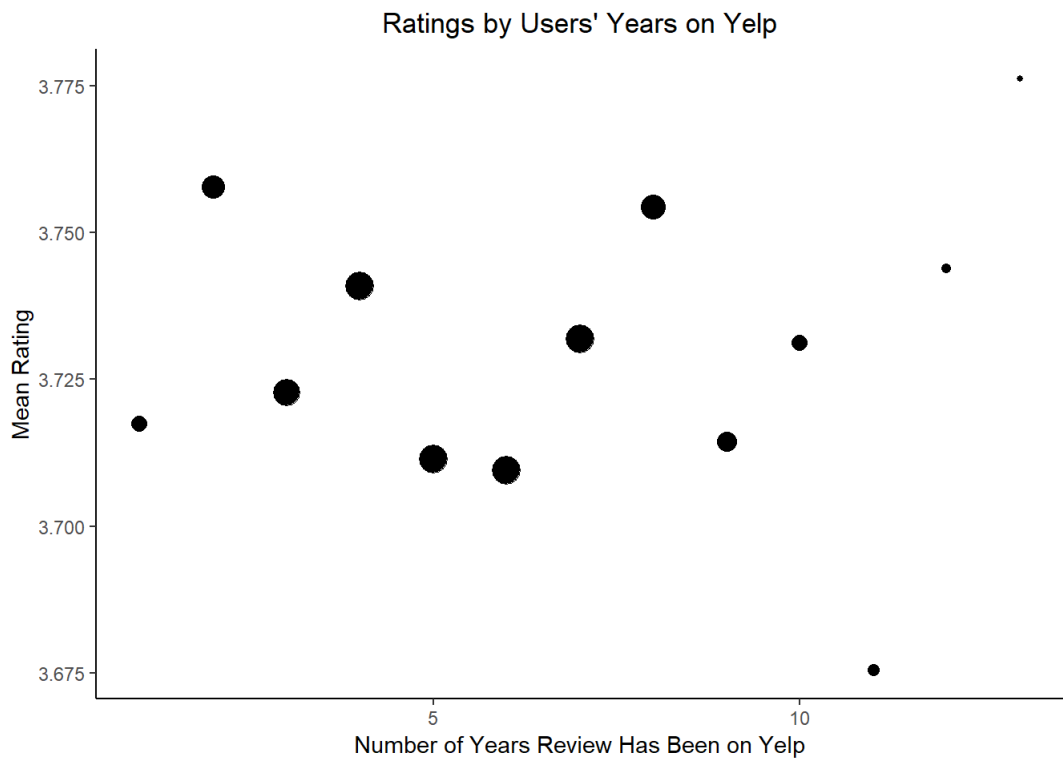
Business Attributes

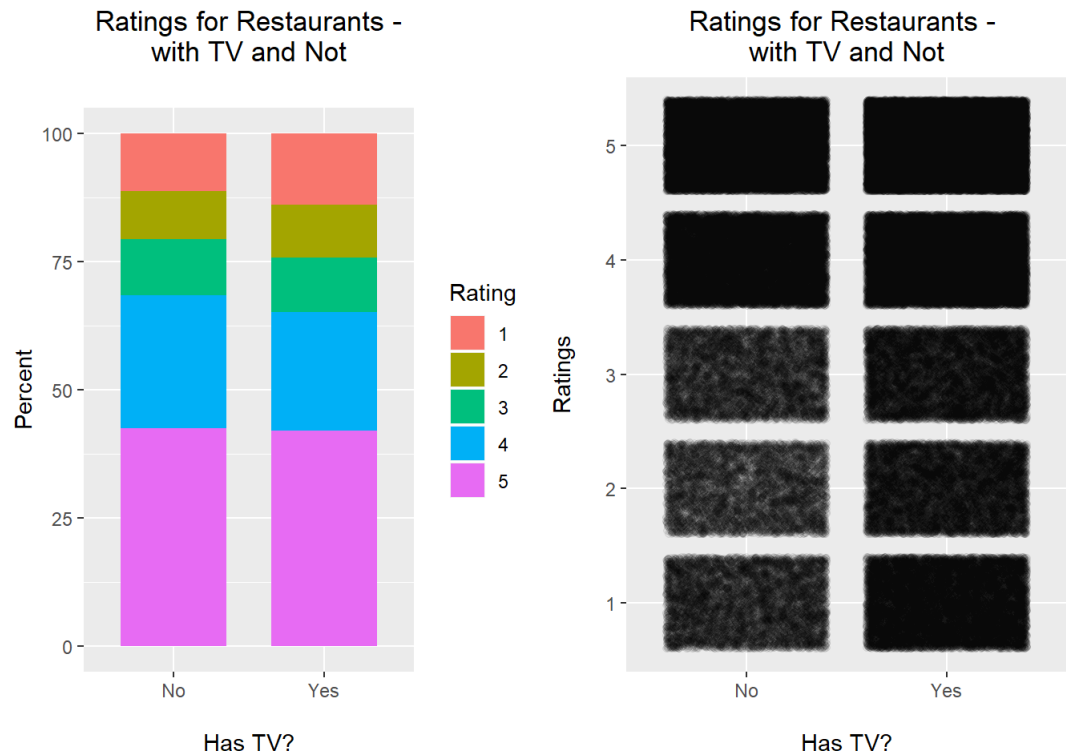Although they don't diverge significantly, restaurants without a TV appear to have receive four or five ratings more frequently than restaurants with at least one TV. There is a large sample size in every category making our jitter plot uninterpretable even at very low alpha levels.

Businesses open 24 hours have slightly more negative reviews. After running the model and seeing such a large standard error for a predictor term based on 24 hour access, I revisited the jitter plot and realized that there are is much smaller sample size for restaurants open 24 hours. I keep this predictor in the model nonetheless while acknowledging its limitations in terms of comparative and predictive value.
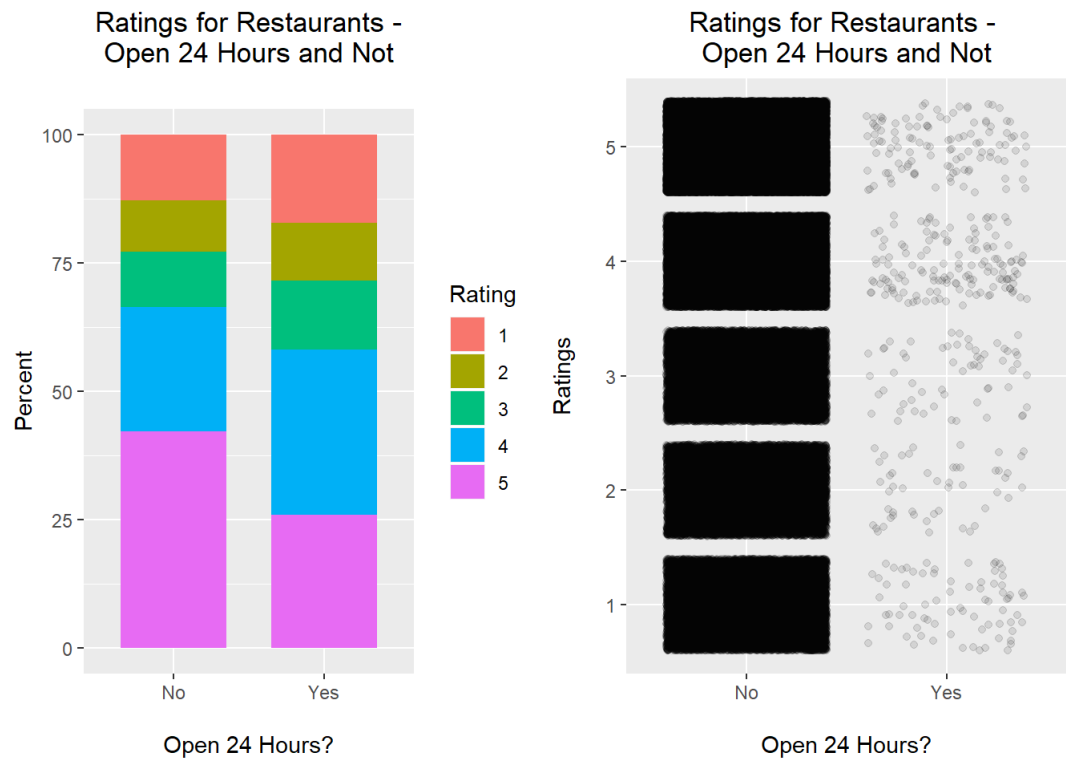
Looking at price cateogires in ascending orders, you can see that restaurants in the lowest price category tend to have the highest proportion of four and five star ratings. After that the proportions vary.

Ratings for Restaurants by Price Category
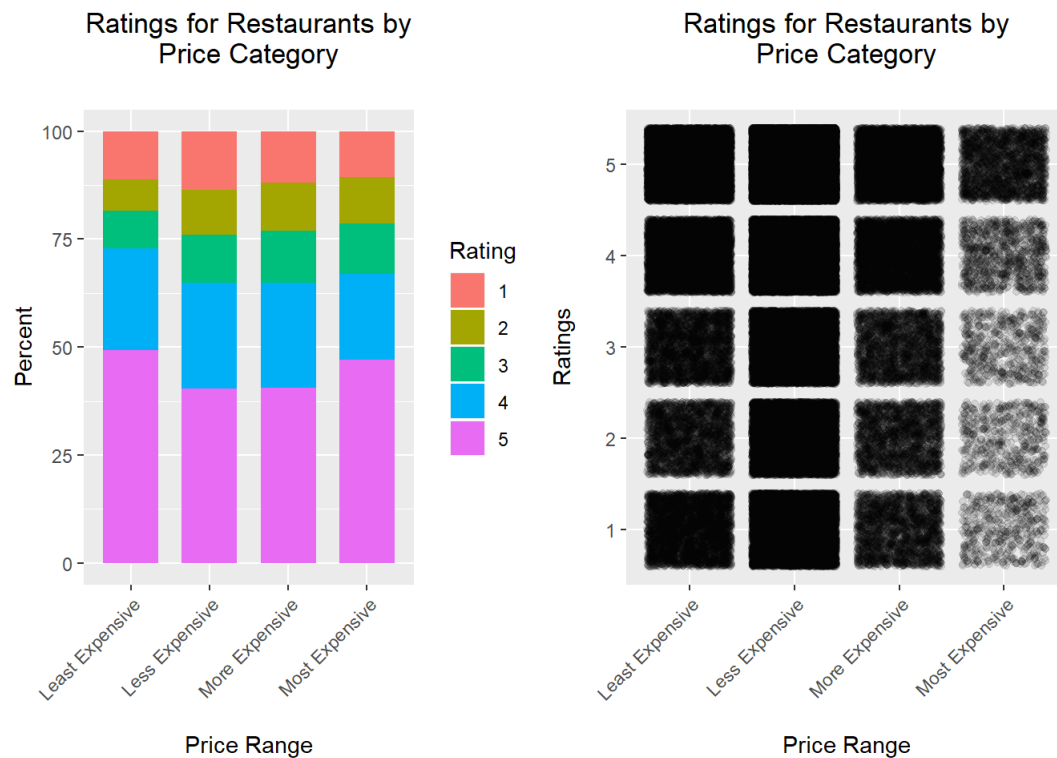
Correlation matrix

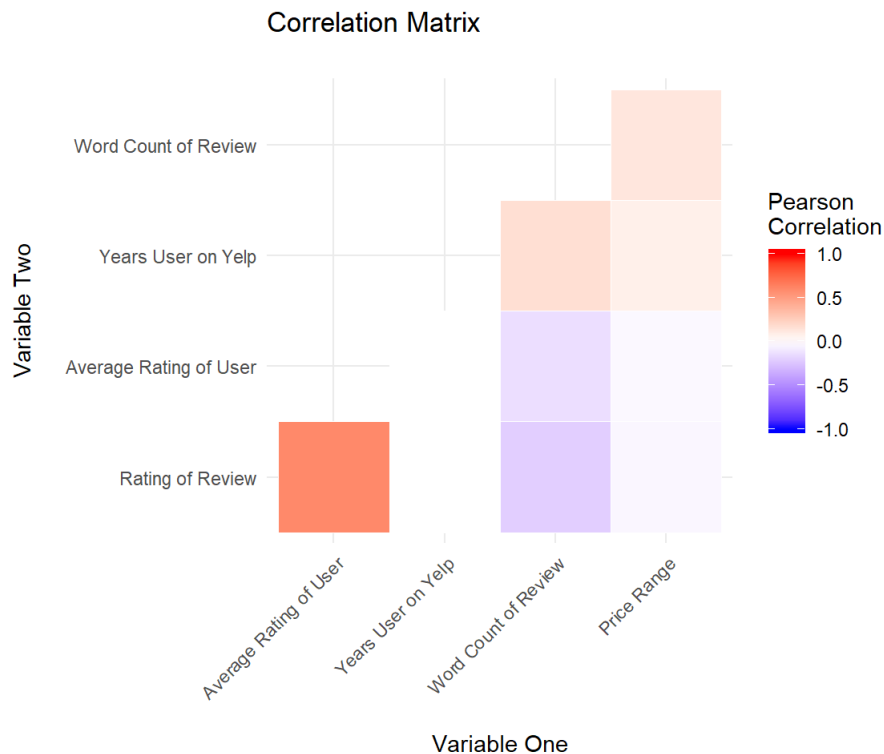Several factors are highly correlated with the rating of an individual review. In particular, longer reviews tend to be more negative, while reviews written by users who write more positive reviews on average tend to be more positive. Reviews of more expensive restaurants are slightly more negative than reviews of cheaper restaurants.

Correlation Matrix

Model Summary - Single Level Ordinal Regression Model

I consider review stars to be an ordinal variable and predict the increased probability of having 2-5 stars, compared to one star. I find that the primary relationships found in the linear regression remain (as seen in appendix). Users with higher average stars are more likely to give higher ratings, and businesses in the second price category are more likely to receive lower ratings. The strongest relationship with ratings is the users' average rating in previous reviews, as one standard deviation's increase in average review stars is associated with a 4 times (e ^ 1.4) increase

in the odds of having one more star when compared to the mean average review stars. Please note that the two continuous variables, mean review stars and years user has been on Yelp, have been centered and scaled while the baseline for the categorical predictors are the lowest price category, no TV, and not open 24 hours a day. The output of the model is presented in graphical, table, and equation form with interpretation following.

```
================================================================================
                              Single Level Ordinal Regression Model
--------------------------------------------------------------------------------
yelping_years_user                              -0.08 ***
                                                (0.02)
average_stars_user                               1.46 ***
                                                (0.02)
factor(RestaurantsPriceRange2_business)2        -0.41 ***
                                                (0.05)
factor(RestaurantsPriceRange2_business)3        -0.28 ***
                                                (0.08)
factor(RestaurantsPriceRange2_business)4        -0.18
                                                (0.13)
factor(HasTV_business)True                      -0.08
                                                (0.04)
factor(Open24Hours_business)True                -0.04
                                                (0.42)
--------------------------------------------------------------------------------
AIC                                          24301.90
BIC                                          24381.22
Log Likelihood                              -12139.95
Deviance                                     24279.90
Num. obs.                                       10000
================================================================================
*** p < 0.001, ** p < 0.01, * p < 0.05
```

Waiting for profiling to be done...

	2.5 %	97.5 %
yelping_years_user	-0.12	-0.04
average_stars_user	1.41	1.50
factor(RestaurantsPriceRange2_business)2	-0.51	-0.30
factor(RestaurantsPriceRange2_business)3	-0.43	-0.13
factor(RestaurantsPriceRange2_business)4	-0.44	0.08
factor(HasTV_business)True	-0.16	0.00
factor(Open24Hours_business)True	-0.86	0.79

Ordinal Logistic Regression Coefficients
(95% interval)
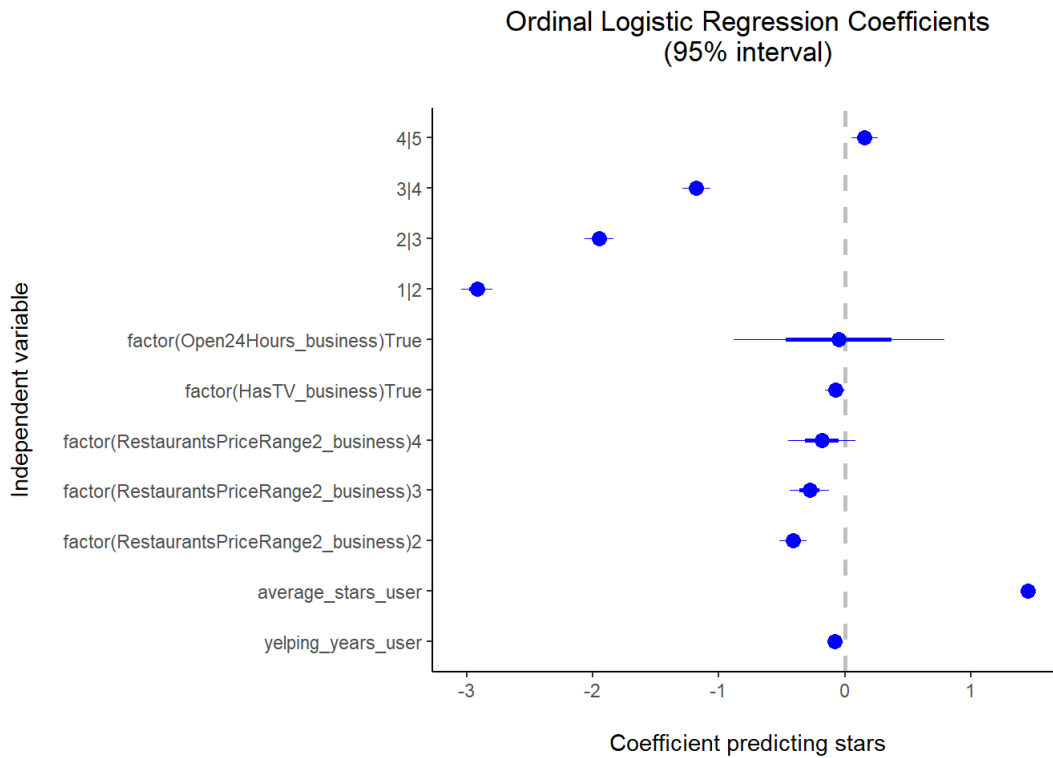
In equation form, our model has generated coefficients for each predictor effect on the cumulative log odds of a rating being a one to five star rating using four threshold points estimated as intercept values. For the probability that any given observation falls in the top and bottom categories, we estimate values by using equations one and four. For equation four we can simply take an inverse-logit of the value on the right side of the equation, while for equation one we can do the same and subtract from one.

For two, three and four star ratings we must take probabilities in adjacent categories and subtract one from another. For example, the probability that a review has two stars can be deduced by taking the inverse logit from equation two, which tells us the probability that the rating is greater than two stars, and subtract the inverse logit from equation one, which tells us the probability that rating is greater than one star. For the sake of compact-ness, we have omitted writing three different indicator terms for price range but acknowledge that is a four level factor.

Equation one:

$$log(\frac{\hat{\pi_2} + \hat{\pi_3} + \hat{\pi_4} + \hat{\pi_5}}{\hat{\pi_1}}) = -\alpha_{1|2} + \beta_1 x_{user.years} + \beta_2 x_{user.avgstar} + \beta_3 x_{price.range} + \beta_4 x_{TV} - \beta_5 x_{open24}$$

Equation two:

$$log(\frac{\hat{\pi_3} + \hat{\pi_4} + \hat{\pi_5}}{\hat{\pi_1} + \hat{\pi_2}}) = -\alpha_{2|3} + \beta_1 x_{user.years} + \beta_2 x_{user.avgstar} + \beta_3 x_{price.range} + \beta_4 x_{TV} - \beta_5 x_{open24}$$

Equation three:

$$log(\frac{\hat{\pi_4} + \hat{\pi_5}}{\hat{\pi_1} + \hat{\pi_2} + \hat{\pi_3}}) = -\alpha_{3|4} + \beta_1 x_{user.years} + \beta_2 x_{user.avgstar} + \beta_3 x_{price.range} + \beta_4 x_{TV} - \beta_5 x_{open24}$$

Equation four:

$$log(\frac{\hat{\pi_5}}{\hat{\pi_1} + \hat{\pi_2} + \hat{\pi_3} + \hat{\pi_4}}) = -\alpha_{4|5} + \beta_1 x_{user.years} + \beta_2 x_{user.avgstar} + \beta_3 x_{price.range} + \beta_4 x_{TV} - \beta_5 x_{open24}$$

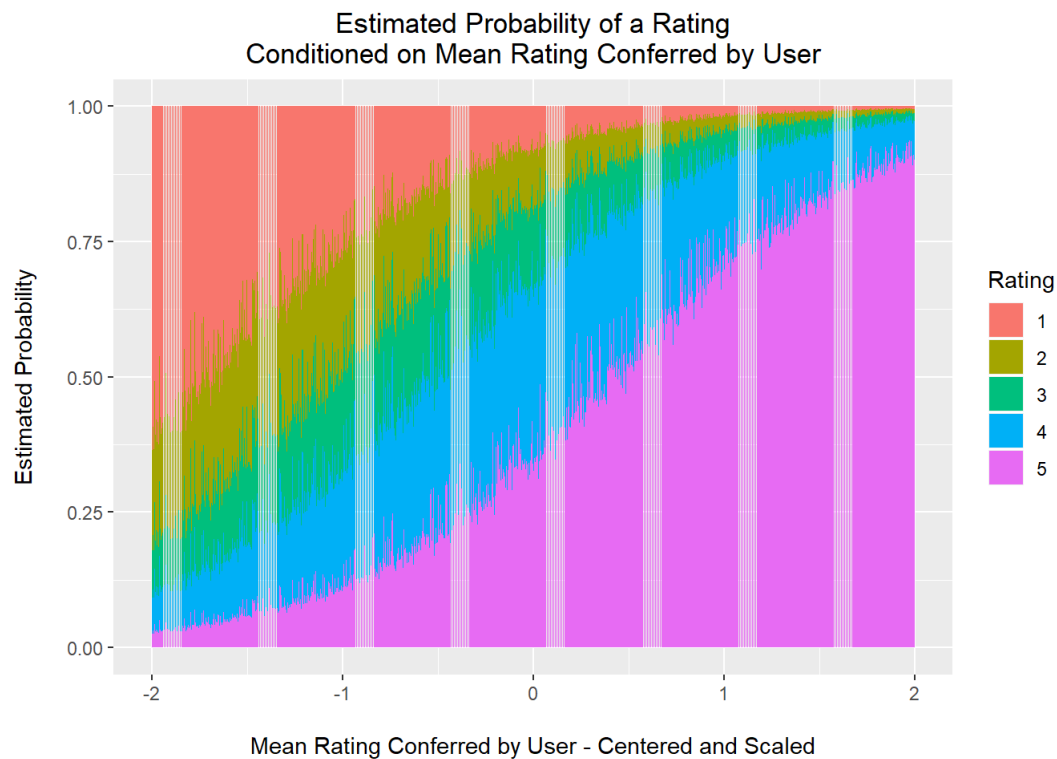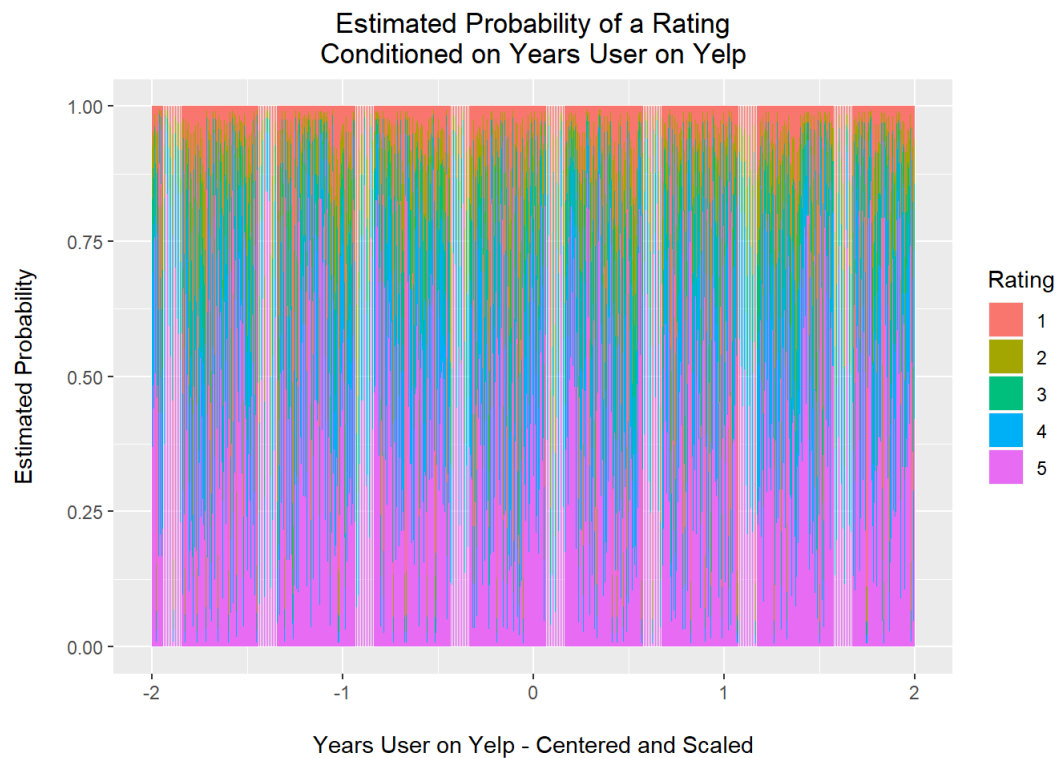# Model visualization - Single Level Ordinal Regression Model

Visualizing the outcome, we plot the predicted probabilities of any given rating against a continuous scale for each continuous predictor, which has already been centered and scaled: the number of years the user has been using Yelp and the average rating the user has conferred overall.

We see a somewhat muddled result for years the user has been using Yelp, but a very clear trend for the user's average rating. The higher the user's average ratings in the past, the higher the probability of giving higher restaurant ratings on any given review.

Estimated Probability of a Rating
Conditioned on Years User on Yelp

Years User on Yelp - Centered and Scaled

Estimated Probability of a Rating
Conditioned on Mean Rating Conferred by User

Mean Rating Conferred by User - Centered and Scaled

Next, I visualize the categorical predictors and expected probabilities holding the two continuous predictors contstant at our centered mean of zero. Our faceted plot grid shows the probabilities within each category for businesses with and without TV within each plot, with and without 24 hour service along the rows, and for price categories along the columns.

You can see roughly that probabilities of higher ratings are greater for businesses without TVs and for businesses without 24 hour service, by comparing top row with bottom row, while acknowledging differences in sample size. Moreover, you see that there is a less clear trend from lowest price category to highest price category, but we see that price category two has the lowest proportions of four and five ratings regardless of any other factors. This seems to directionally agree with out model determining statistical significance of price category two. We keep in mind the miniscule sample size of the 24 hours open restaurants in discounting these results.
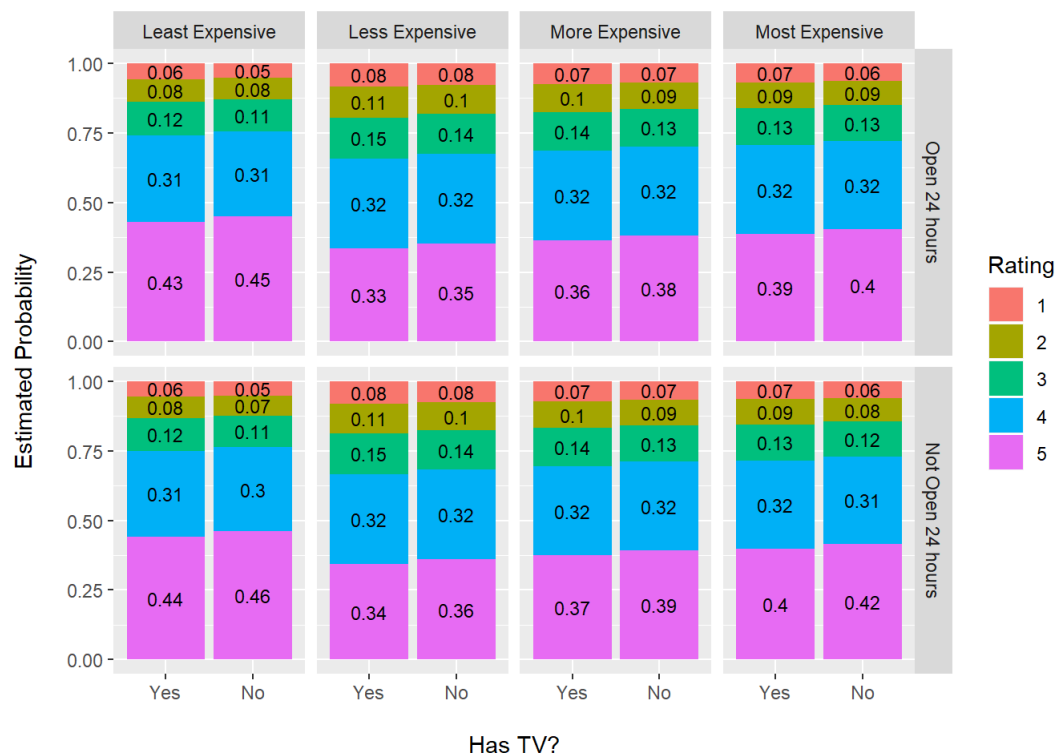
Model Check - Single Level Ordinal Regression Model

Next, I look at a series of residual plots of predictions for each response category. The residuals look reasonable although there appears to be some trend and heteroscedacity; the residuals lie within +/- 0.6 standard units. The binned points are mostly within the two standard error line which may be overconfident. For the five star rating in particular, the binned residuals are a bit more problematic. Here, almost all of the residuals are positive, indicating that the model has underestimated the expected probabilities.
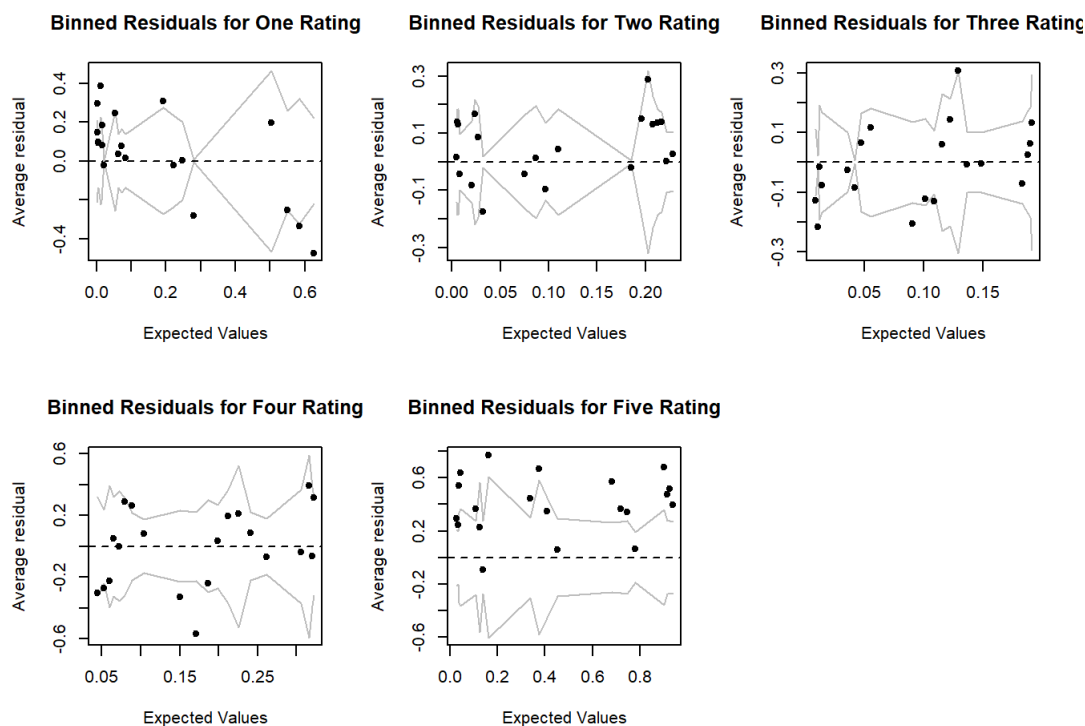
This method of examining multiple residuals corresponding to the ordinal outcomes makes interpretation of this diagnostic tool more challenging. Building off the research of Li and Shepherd (2012), I try using sign-based statistic (SBS), or probabiliy-scale residuals. These residuals are inherently discrete and sometimes display unusual patterns in the diagnostic plots. Li and Shepherd show that the sign-based-statistic (SBS) can be used as a residual for ordinal outcomes. (4) (5)

For an ordinal outcome Y:

$$R_{SBS} = E(sign(y - Y)) = Pr(y > Y) - Pr(y < Y)$$

I plot the probability-scale residuals against two continuous predictors and find that they are unrelated to the years a user has been Yelping, and somewhat related to users' average review stars.

The Q-Q plot indicates that the model looks ok.
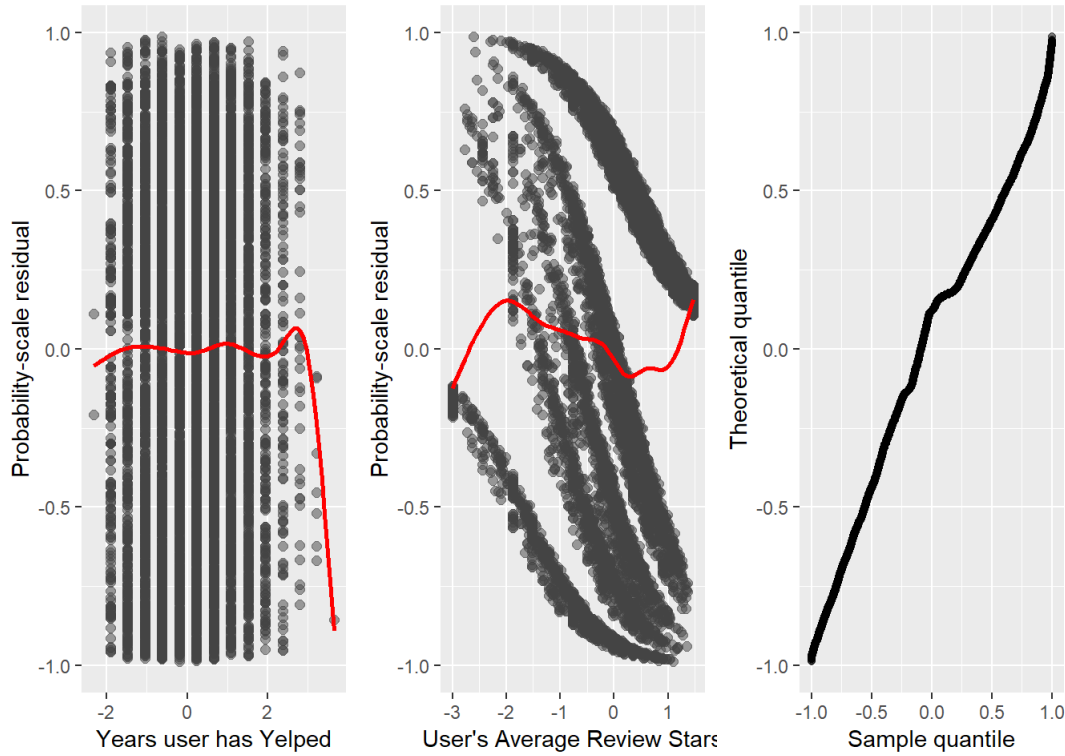
Next, we plot the surrogate residuals which are defined as

$$R_S = S - E(S|X)$$

where S is a continuous variable based on the conditional distribution of the latent variable Z given Y. If Y = y, Liu and Zhang (2017) show that S follows a truncated distribution obtained by truncating the distribution of:

$$Z = -f(X, \beta)$$

using the distribution of:

$$(\alpha_{y-1}, \alpha_y)$$

While acknowledging an incomplete appreciation of the theory's full implications, I leverage a some of the key takeaways that are shown to be true if the model is correct:

1. symmetry around zero
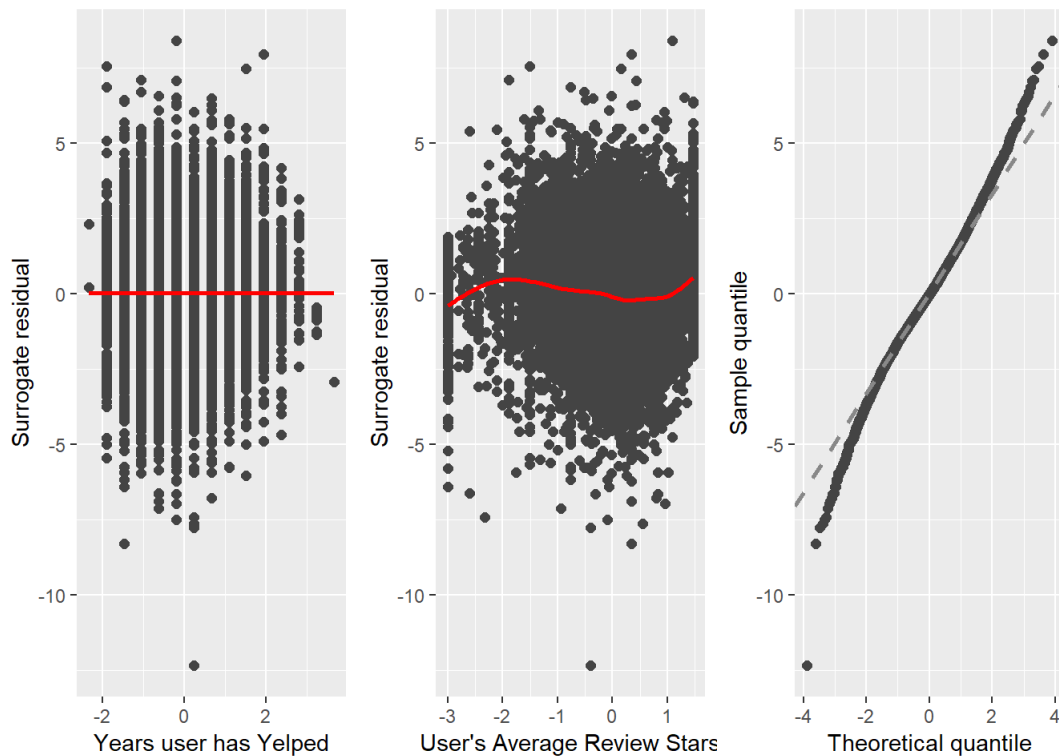
$$E(R_S|X) = 0$$

2. homogeneity

$$Var(R_S|X) = c$$

a constant that is independent of X. (6)

This appears to be mostly satisfied in the plots below, but I note that using this diagnostic, the QQ plot deviates a bit on extreme values of the theoretical quantiles indicating that the model may exhibit heavy tails.

Intraclass correlation coefficient (ICC)

The intraclass correlation coefficient (ICC) is the proportion of variance explained by which business the review is about, rather than the observation-level, which is individual reviews, or random unknown factors. The ICC is quite large, indicating that review stars are highly clustered within businesses. This suggests that wmultilevel model is appropriate to account for this clustering and to avoid overstating confidence about the relationships estimated. The ICC is calculated below as:

```
[1] 0.17
```

The ICC shows that the review observations are not independent of each other, as reviews of the same business are related to one another. For this reason, it is appropriate to fit a multilevel model that accounts for this clustering, while still allowing us to estimate the effect of business-level factors.

Model Summary - Multi-Level Ordinal Regression Model

I fitted a multilevel ordinal logistic regression model with an intercept varying by business and found very similar relationships as the single-level ordinal logistic regression. That is, the most predictive factor remains a user's previous review satisfaction, and restaurants with TVs and restaurants open 24-hours tend to have more negative reviews. This model was fitted using the most commonly discussed method I found from online research: the CLMM function from the R package Ordinal. (7)

This model fits the data better than the single-level model, as the BIC is smaller for the multilevel model. Moreover, compared to the linear and multilevel linear model the BIC remains the lowest.

```
================================================================================
                                Multi-Level Ordinal Regression Model
--------------------------------------------------------------------------------
yelping_years_user                          -0.08 ***
                                            (0.02)
average_stars_user                           1.54 ***
                                            (0.03)
factor(RestaurantsPriceRange2_business)2    -0.45 ***
                                            (0.06)
factor(RestaurantsPriceRange2_business)3    -0.27 **
                                            (0.09)
factor(RestaurantsPriceRange2_business)4    -0.19
                                            (0.16)
factor(HasTV_business)True                  -0.09
                                            (0.05)
factor(Open24Hours_business)True            -0.17
                                            (0.50)
1|2                                         -3.10 ***
                                            (0.07)
2|3                                         -2.08 ***
                                            (0.07)
3|4                                         -1.26 ***
                                            (0.06)
4|5                                          0.16 **
                                            (0.06)
--------------------------------------------------------------------------------
Log Likelihood                          -12103.78
AIC                                      24231.56
BIC                                      24318.09
Num. obs.                                   10000
Groups (business_id)                         5055
Variance: business_id: (Intercept)           0.34
================================================================================
*** p < 0.001, ** p < 0.01, * p < 0.05
```

	2.5 %	97.5 %
1|2	-3.25	-2.96
2|3	-2.21	-1.95
3|4	-1.38	-1.14
4|5	0.05	0.28
yelping_years_user	-0.12	-0.04
average_stars_user	1.48	1.59
factor(RestaurantsPriceRange2_business)2	-0.57	-0.33
factor(RestaurantsPriceRange2_business)3	-0.45	-0.10
factor(RestaurantsPriceRange2_business)4	-0.49	0.12
factor(HasTV_business)True	-0.18	0.00
factor(Open24Hours_business)True	-1.16	0.82

Multilevel Ordinal Logistic Regression Coefficients
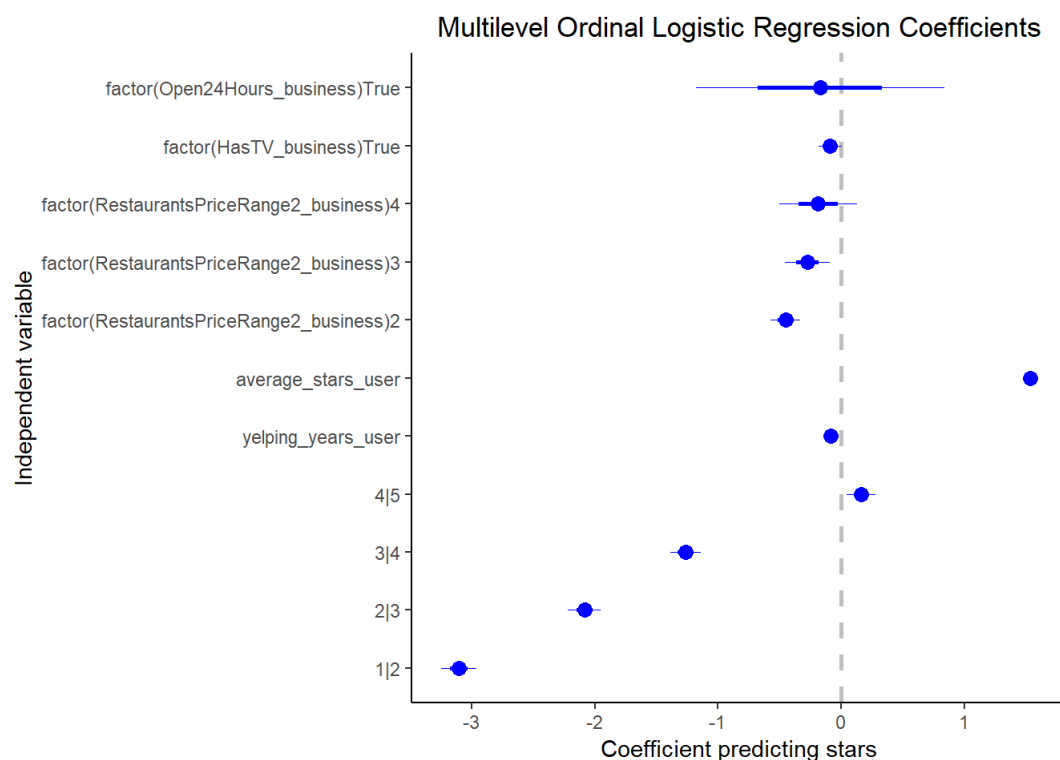
Model	AIC	BIC
Multilevel Ordinal Regression	24232	24318
Single-level Ordinal Regression	24302	24381
Multilevel Linear Regression	30773	30845
Single-level Linear Regression	30811	30876

The output of the multilevel model is similar to the single level regression model except it allows the intercept to vary around its mean value. The model can be generalized as written below - I'll forgo listing and interpreting all four cumulative odds models given its redundance to the four equations written for the single level ordinal regression:

$$log(\frac{Pr(y_i > \pi_i)}{1 - Pr(y_i > \pi_i)}) = \alpha_{j[i]} + \beta_1 x_{user.years} + \beta_2 x_{user.avgstar} + \beta_3 x_{price.range} + \beta_4 x_{TV} - \beta_5 x_{open24}$$

$$\alpha_j \sim N(\mu_\alpha, \sigma^2)$$

Model Checks - Multi-Level Ordinal Regression Model

After extensive research, I omitted examination of residuals and model fit for the multilevel ordinal regression due to technical complications and the absence of clearly defined methods. For the CLMM package used to fit the multi-level ordinal regression, there is not even a compatible predict function. The author of the "Ordinal" package, Rune Christensen, then known as Rune Haubo, commented in response to how to obtain residuals:

"You are not alone in being surprised of the absence of residuals… And I would be delighted to implement residuals for CLMs and CLMMs in ordinal, its only that I have never seen a relevant definition of such residuals or been able to come up with one myself… There is at present no predict method for clmm objects, so here you have another challenge. However, as I indicated above, the real challenge is to figure out what to do with such residuals."

Ben Bolker, a frequently cited expert on mixed effect models, noted on the same thread:

"My first thought is that since fitted(model) works, you might be able to used fitted(model)-observed, but on second thought, you're going to have to figure out what scale the 'fitted' value is on and how it relates to the predicted value of the response."

Given the weight of expert opinion and after an exhaustive search, I conceded that there might not exist a pre-packaged method to apply the same types of analysis and functions to check this model. Nevertheless, the logic stands that a mixed effect model should fit at least as well as the single-level model fit given the intraclass correlations within the same business. This assertion is also reinforced by the fact that the mixed effects model had a lower BIC, even though BIC is relatively conservative and penalizes model complexity more heavily relative than AIC.

Discussion

Both the ordinal regression models agree about the relationships between the factors examined and a positive restaurant experience. In particular, they suggest that restaurants in the second price category (two dollar signs on Yelp) tend to be more negative, while reviews written by Yelpers who usually write positive reviews are much more likely to be positive and those by Yelpers who have been on Yelp for longer are slightly more negative. There is also signifcant clustering within reviews of the same business as evidenced in the ICC.

Together, these findings suggest that when reading a restaurant review, one should think considerably about who is writing the review, as their general outlook may greatly contribute to their review. Furthermore, one should expect that individual ratings of the same restaurant should be close to one another on average.

Limitations

Our study has a multitude of limitations. First we have introduced sampling error even within the Yelp dataset by taking a randome subsample of 10,000. Furthermore, we are unable to do model checking in a systematically robust fashion for the multi-level ordinal regression. In retrospect, I could have also selected more or better predictor variables from the JSON files.

Finally, if we are trying to generalize Yelp reviews to general restaurant experience, we introduce a host of additional limitation including that Yelp! reviews are not a random sample of all restaurant experiences. Rather, users who have particularly interesting, positive, or negative experiences may be more likely to take the effort to write a review. In addition, people who are very busy, concerned about online privacy, or generally do not spend much time online are unlikely to write Yelp! reviews, causing our sample to be unrepresentative. T

Future Directions

Future work could look at more factors related to restaurant experiences, including restaurant cuisine and location, as well as reviewer gender and age. In addition, future software developers and statisticians may also devise more standardized methodologies for examining residuals and adapting similar diagnostic functionality for the class of multi-level ordinal functions as exist for the single level ones.

Reference

Dataset from twelfth annual Yelp Academic Challenge: https://www.yelp.com/dataset (https://www.yelp.com/dataset)

1. "Google, Yelp, Facebook Most Trusted For Online Reviews." 2018. 08/29/2013. Accessed December 9. https://www.mediapost.com/publications/article/328678/google-yelp-facebook-most-trusted-for-online-rev.html (https://www.mediapost.com/publications/article/328678/google-yelp-facebook-most-trusted-for-online-rev.html).

2. Luca, Michael. "Reviews, Reputation, and Revenue: The Case of Yelp.com." Harvard Business School Working Paper, No. 12-016, September 2011. (Revised March 2016. Revise and resubmit at the American Economic Journal - Applied Economics.)

3. Mahaney, Mark. Yelp Inc.: Gaining Speed in the Local Lane. Yelp Inc.: Gaining Speed in the Local Lane. New York, NY: RBC Capital Markets.

4. Greenwell, Brandon, Andrew McCarthy, Bradley Boehmke, and Dungang Liu. 2018. "Residuals and Diagnostics for Binary and Ordinal Regression Models: An Introduction to the Sure Package." The R Journal 10 (1): 381-394. https://journal.r-project.org/archive/2018/RJ-2018-004/RJ-2018-004.pdf (https://journal.r-project.org/archive/2018/RJ-2018-004/RJ-2018-004.pdf).

5. C. Li and B. E. Shepherd. A new residual for ordinal outcomes. Biometrika, 99(2):473-480, 2012. URL http://dx.doi.org/10.1093/biomet/asr073 (http://dx.doi.org/10.1093/biomet/asr073). [p382]

6. D. Liu and H. Zhang. Residuals and diagnostics for ordinal regression models: A surrogate approach. Journal of the American Statistical Association, URL http://dx.doi.org/10.1080/ (http://dx.doi.org/10.1080/)

7. Christensen, Rune HB. Regression Models for Ordinal Data Introducing R-Package Ordinal.

Appendix

Word count vs. stars

Very positive reviews tend to be much shorter than negative reviews.

Review Word Count by Stars
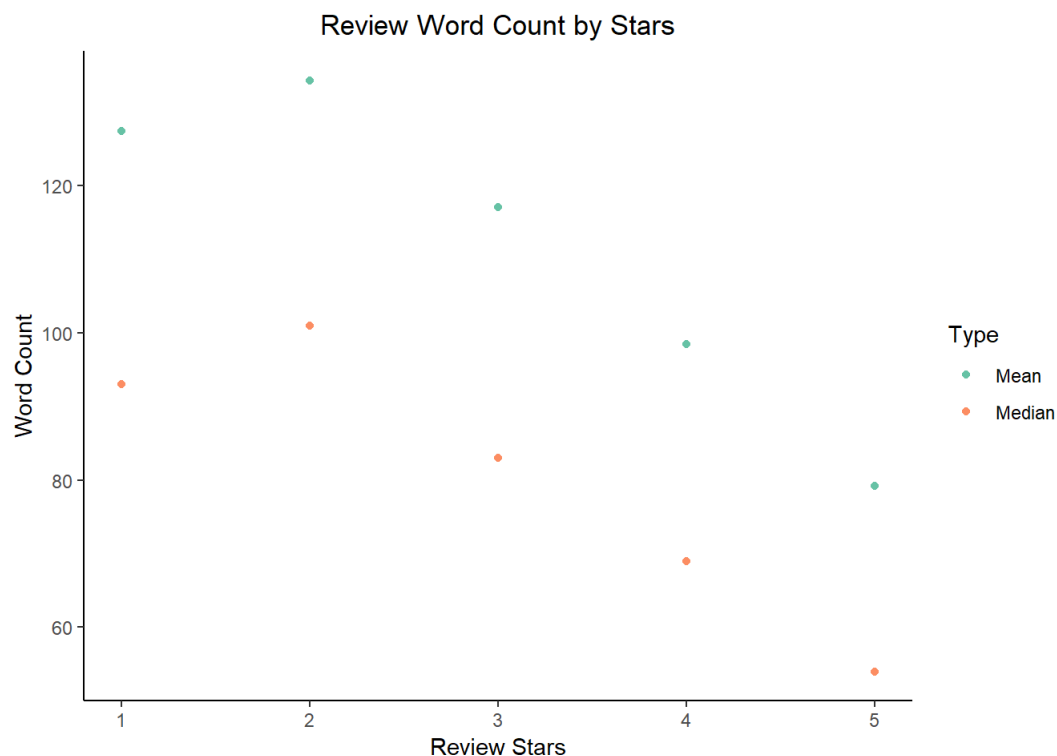
Restaurant review count

WThe following plot shows the relationship between average review ratings (1-5) and the number of reviews written about the restaurant. The blue line is a local regression fit to the points and shows that there is no strong relationship between review count and stars. For this reason we do not include this variable in the regression analyses.

Restaurant review volume by mean stars
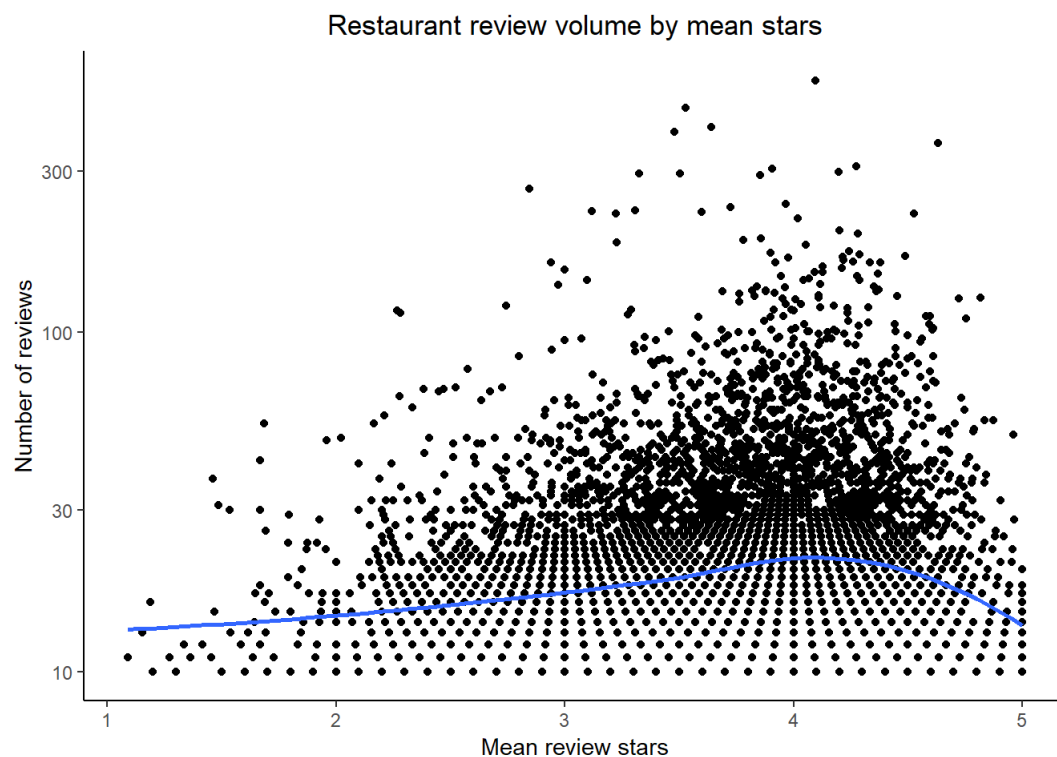
Linear regression

This model suggests that, holding constant the cost of the restaurant, people tend give higher ratings when the restaurant does not have a TV, and particularly when the business is not in the second price category. This makes sense, as such restaurants would likely have lower quality food, and perhaps more frequently have cleanliness and service issues, two common themes in negative reviews.

In addition, this model suggests that the strongest predictor of having a positive restaurant experience is to usually have a positive experience. This may be because certain people are generally more positive than others, or because certain people tend to go to better restaurants and therefore be able to write more positive reviews.

```
================================================================================
                                Single Level Linear Regression Model
--------------------------------------------------------------------------------
(Intercept)                                    3.94 ***
                                              (0.03)
yelping_years_user                            -0.01
                                              (0.01)
average_stars_user                             0.85 ***
                                              (0.01)
factor(RestaurantsPriceRange2_business)2      -0.22 ***
                                              (0.03)
factor(RestaurantsPriceRange2_business)3      -0.14 **
                                              (0.04)
factor(RestaurantsPriceRange2_business)4      -0.11
                                              (0.08)
factor(HasTV_business)True                    -0.06 **
                                              (0.02)
factor(Open24Hours_business)True               0.10
                                              (0.26)
--------------------------------------------------------------------------------
R^2                                            0.37
Adj. R^2                                       0.37
Num. obs.                                   10000
RMSE                                           1.13
================================================================================
*** p < 0.001, ** p < 0.01, * p < 0.05
```

	2.5 %	97.5 %
(Intercept)	3.88	3.99
yelping_years_user	-0.03	0.01
average_stars_user	0.83	0.88
factor(RestaurantsPriceRange2_business)2	-0.28	-0.16
factor(RestaurantsPriceRange2_business)3	-0.23	-0.05
factor(RestaurantsPriceRange2_business)4	-0.26	0.04
factor(HasTV_business)True	-0.11	-0.02
factor(Open24Hours_business)True	-0.41	0.61

Linear Regression Coefficients

Binned residual plot
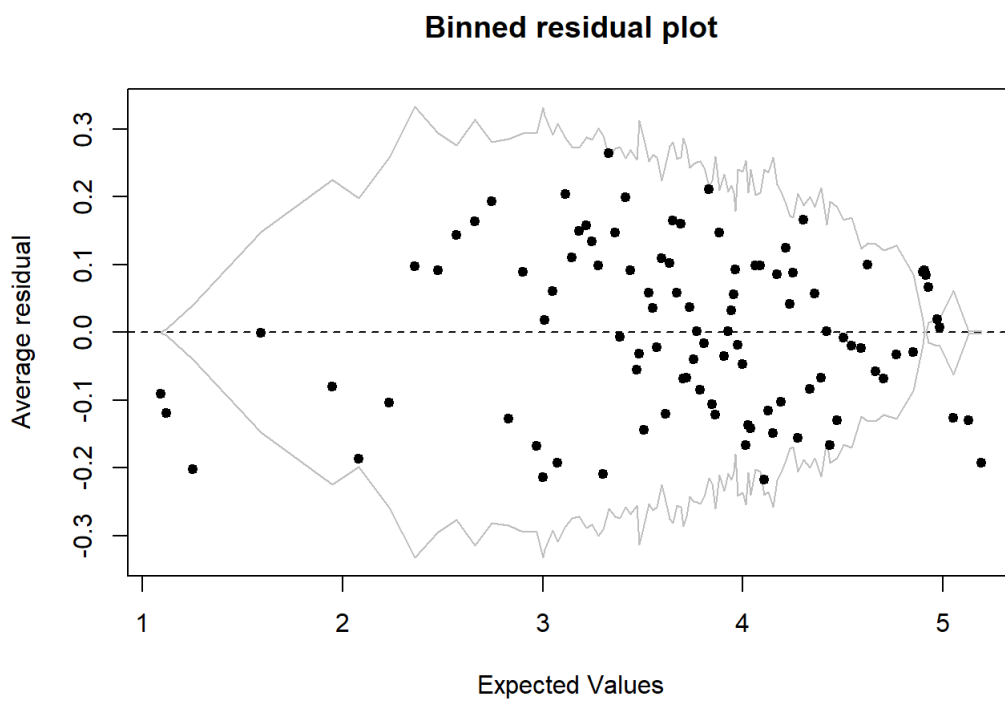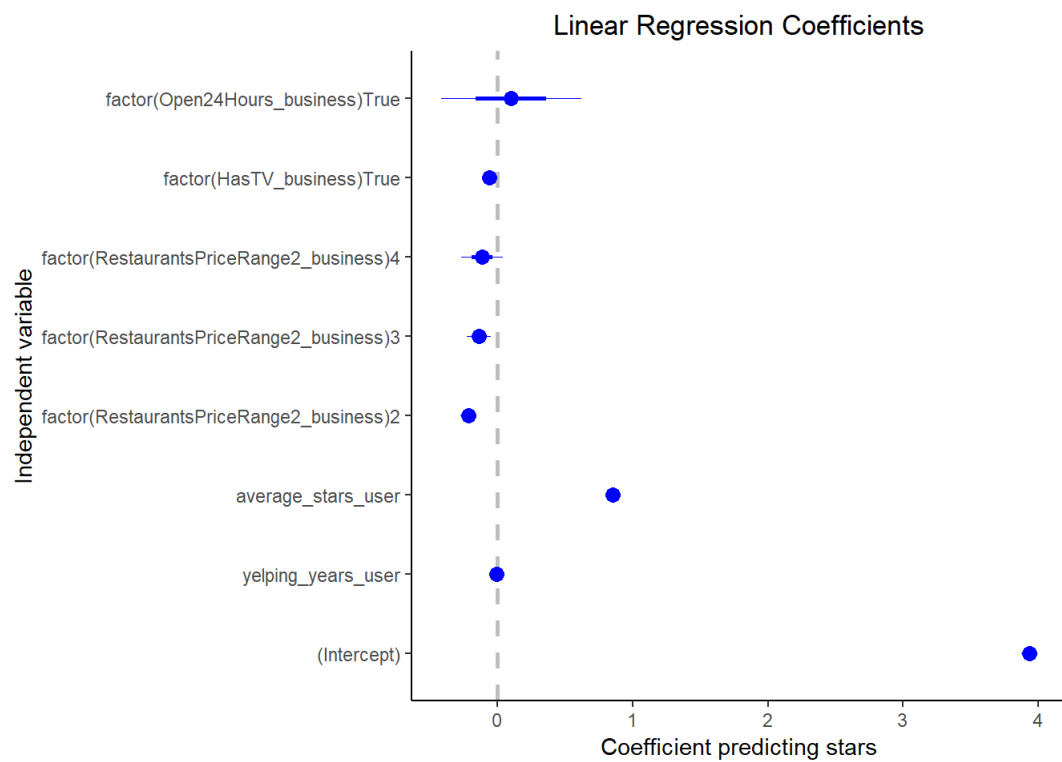
Histogram of predicted
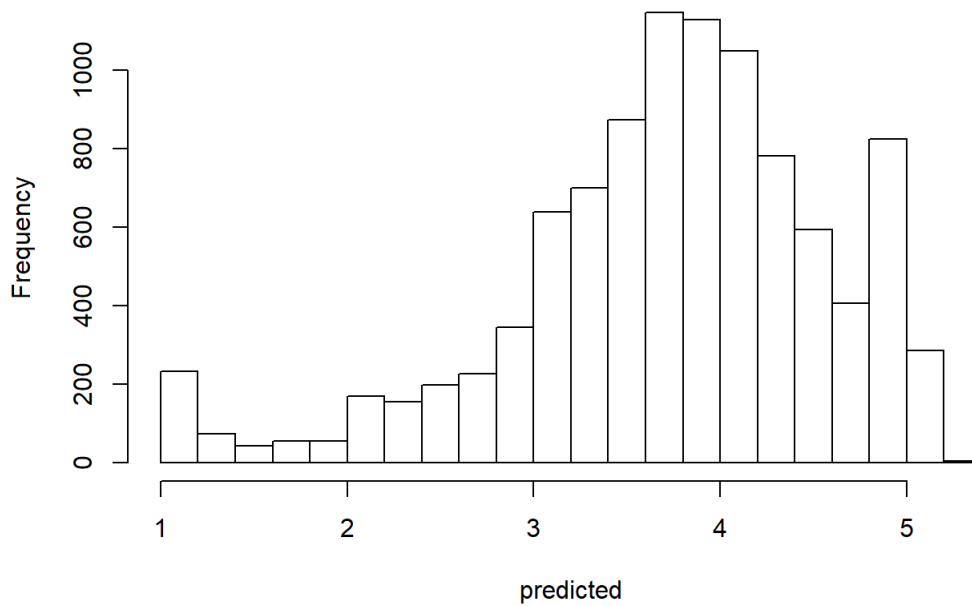
Histogram of df$stars
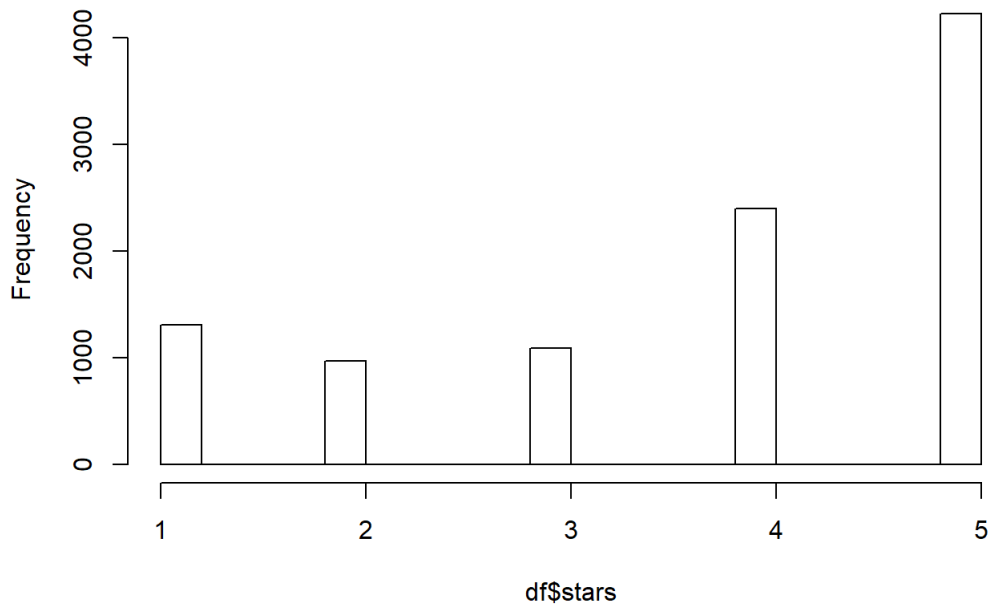
Multilevel linear regression

I fit a multilevel linear regression model with an intercept that varies across businesses. This model allows us to account for clustering at the business level while still being able to estimate the effect of business-level factors.

I find very similar relationships with this model as with the previous ones. The most predictive factor remains a user's previous review satisfaction, and restaurants with TVs and restaurants open 24-hours tend to have more negative reviews.

```
================================================================================
                                    Multi-Level Linear Regression Model
--------------------------------------------------------------------------------
(Intercept)                                 3.93 ***
                                           (0.03)
yelping_years_user                         -0.01
                                           (0.01)
average_stars_user                          0.84 ***
                                           (0.01)
factor(RestaurantsPriceRange2_business)2   -0.22 ***
                                           (0.03)
factor(RestaurantsPriceRange2_business)3   -0.12 *
                                           (0.05)
factor(RestaurantsPriceRange2_business)4   -0.11
                                           (0.09)
factor(HasTV_business)True                 -0.07 **
                                           (0.03)
factor(Open24Hours_business)True           -0.00
                                           (0.30)
--------------------------------------------------------------------------------
AIC                                      30772.97
BIC                                      30845.07
Log Likelihood                          -15376.48
Num. obs.                                  10000
Num. groups: business_id                    5055
Var: business_id (Intercept)                0.12
Var: Residual                               1.15
================================================================================
*** p < 0.001, ** p < 0.01, * p < 0.05
```

```
Computing profile confidence intervals ...
```

	2.5 %	97.5 %
.sig01	0.30	0.39
.sigma	1.06	1.09
(Intercept)	3.87	3.99
yelping_years_user	-0.03	0.02
average_stars_user	0.82	0.86
factor(RestaurantsPriceRange2_business)2	-0.29	-0.16
factor(RestaurantsPriceRange2_business)3	-0.21	-0.02
factor(RestaurantsPriceRange2_business)4	-0.28	0.06
factor(HasTV_business)True	-0.12	-0.02
factor(Open24Hours_business)True	-0.59	0.58

Multi-Level Linear Regression Coefficients

Binned residual plot