MIDTERM PROJECT FOR MATH 156/E-156
Chang Liu and Jennifer Le Hegaret
December 7, 2014

yelp Analytics – Tasty!

Introduction

Chang's personal inspiration for doing a Yelp analytics project comes from his college year, where he was stuck in where there was little access to his favorite: Chinese food. Struggling to find out which restaurant to go to when there was finally a chance, he was obsessed with Yelp; knowing more statistics, he was inspired to dig into the business secrets of Yelp – reviews and rating – and find out more, hence the project.

Data

From the Yelp API[1], we downloaded 999 restaurant data in the Boston Metro Area:

- **average rating across all reviews**
- **number of reviews**
- **restaurant is "claimed" by the owner?**
- **location by zipcode and neighborhood**
- **type of food (Chinese, Vegan, Dive Bar, etc)**

From the US Census[2], we added this information by zip code: median family income, overall population, % of aged 20-40 years old, and % of Asia population.

What questions are we most curious about?

- When does a restaurant know it's got enough "buzz"? *
- Can the ratings be trusted? *
- How good is Boston food, anyway?
- Can it do anything about that by getting "claimed"?
- Do more people in an area lead to more restaurants? Or just certain people?
- Is Chinatown the only place to find Chinese food?

The Answers (and the Surprises!)

We innovatively use Bayesian techniques to estimate the median number of restaurant reviews in the Boston Metro area to be somewhere between 129 and 150. Bootstrapping our sample of 1000 restaurants backs this up.

Is there any known distribution that could have generated the ratings? We find out that we can model the ratings using a normal distribution due to the CLT effect! And the bootstrap student t confidence interval tells us the rating for Boston is 4.

From the question above, we stumble upon a discovery that the ratings and the number of reviews for a restaurant have a negative but weak correlation.

Surprisingly, linear regression shows us that overall, more people in an area is not correlated with a greater number of restaurants in the area. However, a higher percentage of people aged 20-40 in the area is correlated with an increase in the number of restaurants in that area.

Are there more Chinese restaurants where there are more Asian? Sure, if you count Chinatown. But, outside of Chinatown, the correlation between the two is only moderate.

---

[1] Available at http://www.yelp.com/developers/

[2] Available at http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml