Report of Dynamic Pricing Model Project

This is the dynamic pricing project I had been working on during summer intern in Redphare. Given basic information of an apartment, it could automatically knit a pdf report within minutes. The length of the report is strictly limited to 1 page; everything is dynamic, except for figure titles. I attached reports of 2 random apartments in Boston in the following pages, they were knitted with R Markdown.

The first section describes your apartment, tells you how many competitors are there in this area, and what is your ranking among them in terms of prices. It also gives you a suggested price, with 95% confidence interval.

Then it plots some charts, showing you price distributions in your neighborhood; your and your competitors' locations, price trend within 4 months in this area, and why you might want to turn your living room into a bedroom (as number of beds influences price per bed in many cases).

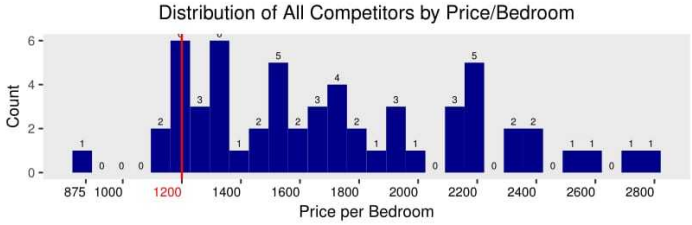
In the last part, it presents 3 competitors, with descriptions and pictures; so that you could virtually compare your department with them, and hopefully have a better understanding of the ranking and pricing of your apartment.

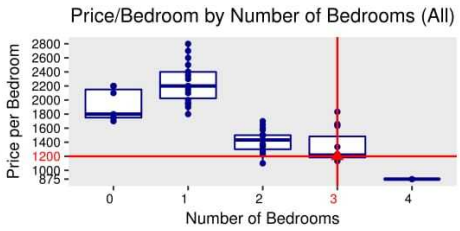
Some facts worth talking about:
1. When looking for competitors, nearby train stations were taken into consideration: rentals within 5 minutes' walk of your apartment would be the first layer competitors; rentals within 5 minutes' walk of your nearby train stations, will be your second layer competitors. The union of them are defined as total competitors.

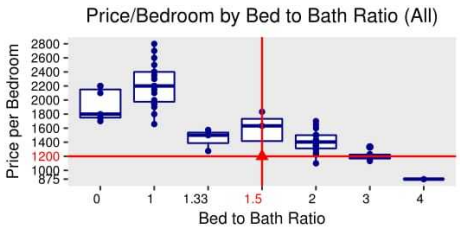
2. The reason why we use only competitors' data in this pricing model: it is generally accepted that location is the key, when it comes to pricing an apartment.

3. This dynamic pricing model greatly increases efficiency of rental agents in Redphare.

15 Highland Avenue, Cambridge, 2017-09-01 , 3bd, 2ba, listed price: $3600, 0 nearby T stations
57 competitors in total, 52 have higher price, 4 have lower price, 1 have same price.
4 direct competitors, 2 have higher price, 1 have lower price, 1 have same price.
(Direct competitors are those with same number of bedrooms and same available date.)
Suggested rent is $3611 per month , with 95% confidence interval: from $3529 to $3693 .







Direct competitor with lowest price.

Update date: 2017-04-23 13:03:00 Type: CBF
Rent: 3500 , 100 lower than
client. 226 meters away from client.
ID: 2378015 Location: Chatham St., Cambridge
(Mid Cambridge) Rent: $3500 Month
Broker Fee: One Month Available Date:
09 01 2017 Beds: 3 Baths:
1 Pet: Cat Ok Features: Balcony,
Coin-op Laundry, Dishwasher, Eat-in Kitchen, Gas
Range, Hardwood Floors, High Ceiling, Kitchen
Pantry, Laundry Facilities, Laundry in Building,
Natural Light, Pantry, Porch, Range, Refrigerato

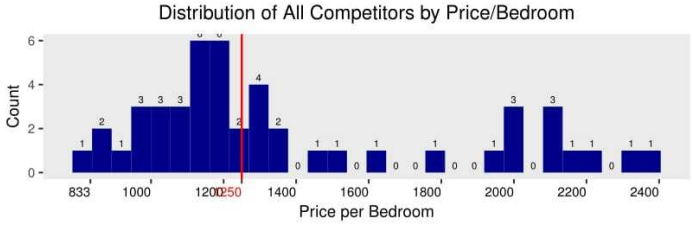
Direct competitor closest to client.

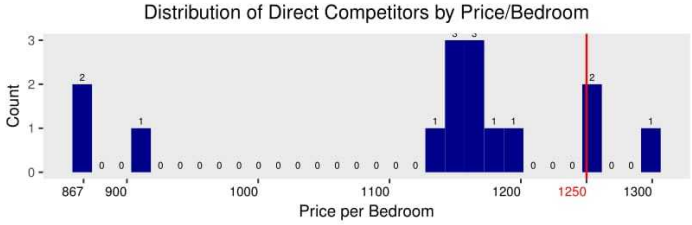
Update date: 2017-04-29 18:38:00 Type: MLS
ID Rent: 3600 , 0 lower
than client. 288 meters away from
client. HIGHLY DESIRABLE AREA BETWEEN INMAN
AND HARVARD SQUARE! Unit boast three
spacious bedrooms, one bath and large
living room. Professionally managed building featuring
common laundry in basement and bike
racks on common patio. Parking available
for rent. Unit is less than
one mile from Harvard Square! Easy
access to shops, restaurants, and Whole

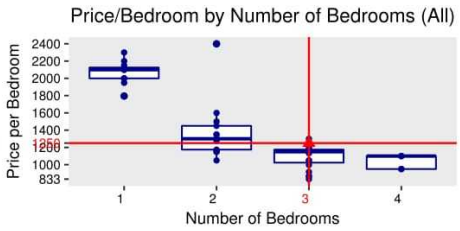
Direct competitor with highest price.

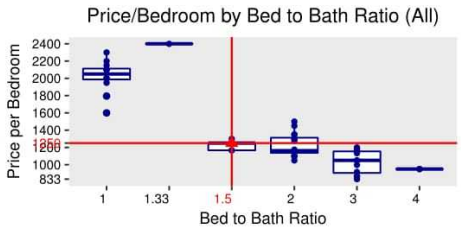
Update date: 2017-04-25 22:19:00 Type: CBO
Rent: 3695 , 95 higher than
client. 370 meters away from client.
Contact info: Eric Sorenson | Micozzi
Management | show contact info |
show contact info Amazing Cambridge 3
bed! Kitchen for miles, porch, ht
hw inc *NO FEE* Sept 1*
39 Clinton St 1-MM, Cambridge, MA
02139 $3,695 mo KEY FEATURES Bedrooms:
3 Beds Bathrooms: 1 Bath Lease
Duration: 1 Year (See Details Below)

10 Hamlin Street, Cambridge, 2017-09-01 , 3bd, 2ba, listed price: $3750, 0 nearby T stations
47 competitors in total, 28 have higher price, 18 have lower price, 1 have same price.
14 direct competitors, 4 have higher price, 9 have lower price, 1 have same price.
(Direct competitors are those with same number of bedrooms and same available date.)
Suggested rent is $3357 per month , with 95% confidence interval: from $3144 to $3570 .









Direct competitor with lowest price.

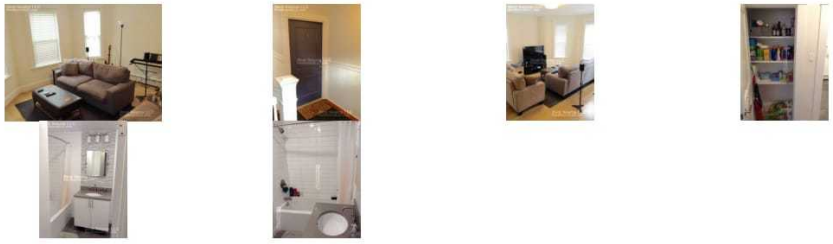
Update date: 2017-04-29 18:39:00 Type: MLS
ID Rent: 2600 , 1150 lower
than client. 343 meters away from
client. Renovated small 3 bedroom unit
located steps to Inman Square. This
unit features gleaming hardwood floors, an
EIK with granite countertops and stainless
steel appliances, and a tiled bathroom
including a fully tiled shower. Coin
op laundry in the building. The
price includes heat and hot water.
No dogs.

Direct competitor closest to client.

Update date: 2017-04-24 11:44:00 Type: CBO
Rent: 3750 , 0 lower than
client. 182 meters away from client.
Apartment Description: Available : AUGUST 1st.
NEWLY RENOVATED -Great Location near shopping
and Great Restaurants-- walk to Kendall
T -Only .2 Mile to One
Kendall Sq. -Large Renovated Eat-In-Kitchen -2
Full Renovated Bathrooms -Nice Back yard
-NO Pets -Free In-Building Laundry -Please
Only Text show contact info BROKERS
WELCOME TO TEXT

Direct competitor with highest price.

Update date: 2017-04-23 21:31:00 Type: CBF
Rent: 3900 , 150 higher than
client. 312 meters away from client.
ID: 2386508 Location: Columbia St., CambridgeRent:
$3900 Month Broker Fee: One
Month Available Date: 09 01 2017
Beds: 3 Baths: 2 Pet: Cat
Ok Rent Includes: Hot Water Super
easy commute to Central Sq -
right on Columbia St. 1st floor,
newly renovated 3 bed with sizable
eat in kitchen, laundry in unit,



Report of BU Student Mental Health Project

Introduction

The client of this project is Boston University Student Health Service Office; mental health survey results of more than 2500 BU students are provided by them. There are 4 main requests from the client. First of all, the team is requested to present the drug use behaviors of Boston University students. Secondly, anxiety level, depression level and happiness level distributions of BU students are required as well. Thirdly, BU Student Health Service Office would like to assess students' satisfaction levels of their consulting service, and find out the demographic features of unsatisfied students. In addition, they are curious about the students who should have mental health consulting while have not experienced any, so that the office could target at certain groups of students and assist them better.

Materials

Materials used in this project are Excel sheets (containing raw data) and corresponding code book provided by clients; the majority of visualization and analysis have been done in R; results and discussion are presented through Excel and PowerPoint.

Methods

Bar plots are used to show the drug usage of BU students. All the observations are divided into several groups, by gender (3 groups), race (6 groups), year of school(6 groups) and major (3 groups) respectively. Bar charts for each grouping method are plotted. Proportion tests are carried out to test whether differences among groups are significant. Density charts and histogram are plotted to show the overall picture of BU students mental health status. The third question is solved with a ordinal regression model, with likert scale (from very dissatisfied to very satisfied) as response, demographic features as predictors. To identify those who should have mental health consulting while have not (students in "help gap"), flow

charts are plotted for assistance; also, after identification of “help gap” group, t tests are conducted to test whether or not the difference of certain variables between this group and “help covered” group are significant.

Results

Visualization of BU students’ drug using behaviors shows that most of BU students are not addictive to harmful drugs, though a certain proportion of them have tried recreational drugs. While mental health status plots could confirm the hypothesis that most of BU students are mentally healthy. As to the third question, ordinal regression results indicate that, several groups of students are significantly less satisfied than the others with some specific aspects of the mental health service; students in “Help gap” are identified, and reasons for their not reaching out for help turn out to be financial status and unawareness of the service.

Discussion

Charts and corresponding statistical tests of question 1 and 2 could present the client with a general yet clear and concise picture of BU students’ status. Question 3 and 4 results might become motivations for BU Mental Health Office to adjust their work strategy to certain extend.

Conclusion

In conclusion, BU students are generally mentally healthy; only a very small proportion of them used to try harmful drugs. Those unsatisfied with mental health service are identified and reasons for dissatisfaction have been found out; demographic features of those need help are also presented to the client, together with reasons for not reaching out.

Report of Children Vocalization Project

Introduction

The client of this project is Dr. Jill Thorson from Boston University Psychological and Brain Sciences Department. Experimental study data of voice pitch type of 12 children and 6 adults in 3 types of conditions(labeled as “New”, “Given” and “Contrastive”) is provided by them. The main request of clients is to compare the voice pitch pattern between children and adults controlling for condition.

Materials

Materials used in this project are Excel sheets (containing raw data) and corresponding code book provided by clients; the majority of visualization and analysis have been done in R; results and discussion are presented through R Markdown and PDF files. After data cleaning, the dataset consists of 5 important variables. The Group variable is whether the observation was initiated by a kid or an adult. The ID variable documented the person who gave the response. Item variable is the item said by the person. Category variable is the type of scenario the items were presented and the PitchTypeResponse variable is the type of response given by the person.

Group	ID	Item	Category	PitchTypeResponse	H-Match	Deacc-Match	LH-Match	group	category	pitch
0	12	11	2	1	1	0	0	child	contrastive	H
0	12	1	1	3	0	0	0	child	given	L
0	12	6	2	3	0	0	0	child	contrastive	L
0	12	9	2	2	0	0	1	child	contrastive	LH
0	12	13	2	1	1	0	0	child	contrastive	H
1	13	0	0	1	1	0	0	adult	new	H
1	13	3	0	1	1	0	0	adult	new	H
1	13	0	1	0	0	1	0	adult	given	D
1	13	3	2	3	0	0	0	adult	contrastive	L

Methods

The core method we use is Mixed-Effects Logistic Regression, controlling for conditions. Since there are 3 types of unordered voice pitch, we fit 3 separate logistic models with random effects (random intercept) to model the probability of giving certain response (“H”, “D”, or “LH”). In general, all three models can be expressed as:

$$log\left(\frac{p}{1-p}\right) = \overline{\beta_0} + \beta_1 x_{Groupchild} + \beta_2 x_{Categorygiven} + \beta_3 x_{Categorynew} + \beta_4 x_{Groupchild:Categorygiven}$$

$$+ \beta_5 x_{Categorychild:Categorynew}$$

Then, for each of the three models, we conducted a generalized linear hypothesis test integrating multiple tests for all three categories where corrections of p-value were involved, the idea is to test whether there is a difference in a certain response (“H”, “D”, “LH”) for a certain category (“new”, “given”, “contrastive”). For each model, the three null hypotheses we are making can be expressed as:

$$Constractive : H_0 : \beta_0 + \beta_1 = \beta_1$$

$$Given : H_0 : \beta_0 + \beta_1 + \beta_2 + \beta_4 = \beta_0 + \beta_2$$

$$New : H_0 : \beta_0 + \beta_1 + \beta_3 + \beta_5 = \beta_0 + \beta_3$$

Or, equivalently:

$$Constractive : H_0 : \beta_0 = 0$$

$$Given : H_0 : \beta_1 + \beta_4 = 0$$

$$New : H_0 : \beta_1 + \beta_5 = 0$$

Results

The test results for each of the three models are listed below.

null.model1.H	estimate	std.error	z.value	p.value
groupchild + groupchild:categorygiven	0.9991	0.4864	2.054	0.107
groupchild	0.7900	0.5078	1.556	0.294
groupchild + groupchild:categorynew	0.4862	0.4745	1.025	0.631

null.model2.D	estimate	std.error	z.value	p.value
groupchild + groupchild:categorygiven	-0.6675	0.4064	-1.643	0.272
groupchild	-0.1497	1.2491	-0.120	0.999
groupchild + groupchild:categorynew	1.8967	1.1029	1.720	0.235

null.model3.LH	estimate	std.error	z.value	p.value
groupchild + groupchild:categorygiven	-0.7318	0.8455	-0.866	0.736
groupchild	-0.4487	0.6181	-0.726	0.824
groupchild + groupchild:categorynew	-0.9998	0.6101	-1.639	0.250

Discussion

For each regression, 3 null hypothesis tests are conducted: for “new” category, “given” category and “contrastive” category respectively. To interpret this result, we use 0.107 as an example. 0.107 indicates that given the null hypothesis is true, ie with fixed category “given”, children and adults have the same probability to use “H” pitch accent, the probability of observing such data or more extreme data is 10.7%.

Conclusion

We can not reject any of the null hypothesis at 5% or 10% significance level. Though the p value of 10.7% suggests that children might have more tendency to use “H” pitch accent than adults, with the “given” category.

Report of Gender Equivalence Project

Introduction

The client of this project is Professor Deborah Belle and her research team, from Boston University Psychological & Brain Sciences Department. Experimental study data of 38 types of children's behaviors, across 20 countries is provided by them. The main request of clients is to assess the correlation between the GGI(Gender Gap Index) and the gender equality levels of these countries. Gender equality levels would be measured by 38 types of behaviors of boys and girls.

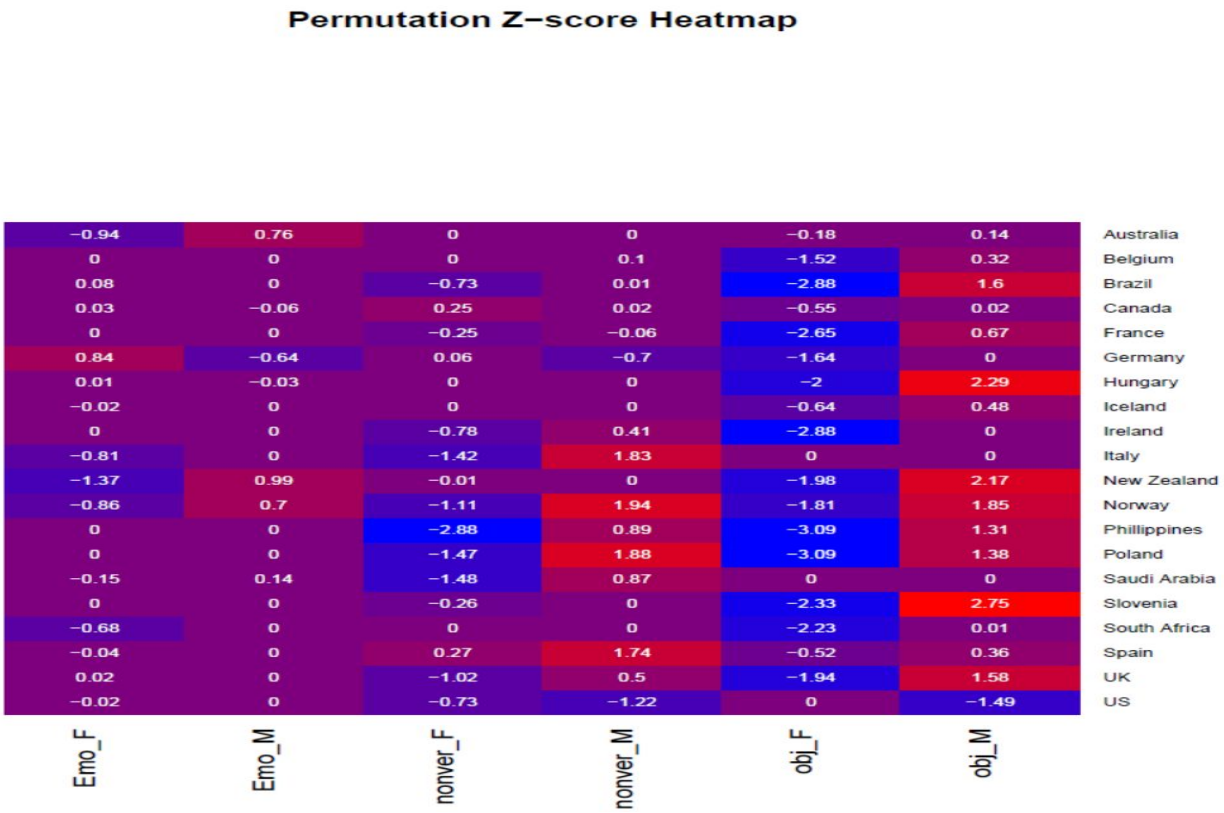
Materials

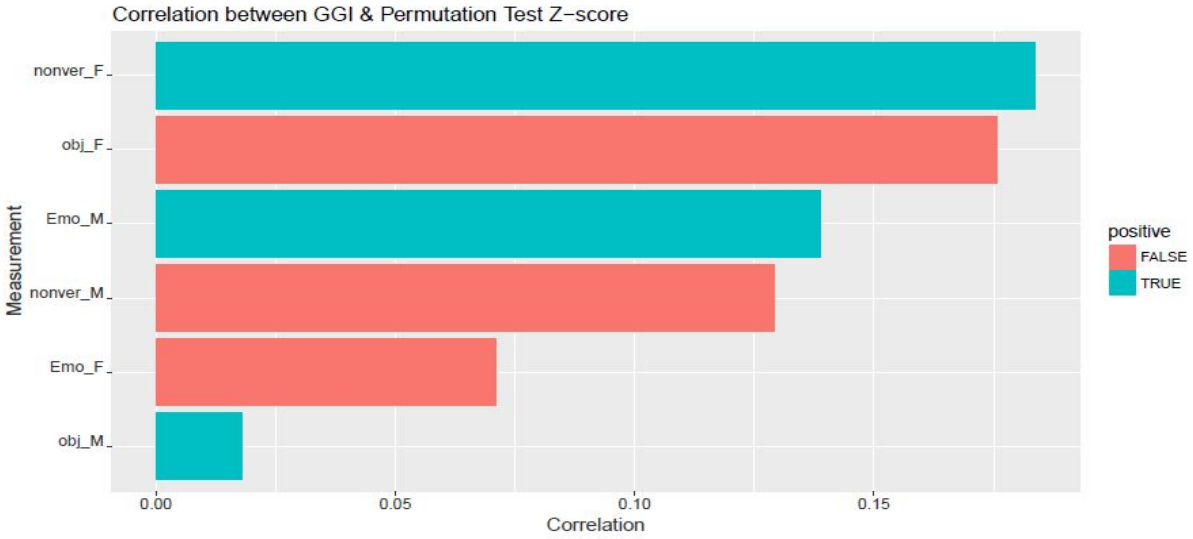
Materials used in this project are Excel sheets (containing raw data) and corresponding code book provided by clients; the majority of visualization and analysis have been done in R; results and discussion are presented through R Markdown and PDF files.

Methods

38 measurements were recategorized into 6 composite variables, after discussions with clients. After which a table of proportions, for each gender, each country, and each composite variable is computed; each cell of this table contains the percentage of girls (or boys) of one of the 20 countries who was observed to have one of 6 behaviors described by composite variables. Then, permutation test p values between genders for each country and each composite variable are calculated. These p values have different degrees of freedom due to different sample size, so they are converted into standard normal distribution scale by inverse probability transformation, for more convenient illustration. Finally, heat map of p values is plotted to show the differences of 6 composite variables among countries; bar charts are used to illustrate the correlation between GGI and composite variables.

Results


Permutation Z−score Heatmap
−0.94 0.76 0 0 −0.18 0.14 Australia
0 0 0 0.1 −1.52 0.32 Belgium
0.08 0 −0.73 0.01 −2.88 1.6 Brazil
0.03 −0.06 0.25 0.02 −0.55 0.02 Canada
0 0 −0.25 −0.06 −2.65 0.67 France
0.84 −0.64 0.06 −0.7 −1.64 0 Germany
0.01 −0.03 0 0 −2 2.29 Hungary
−0.02 0 0 0 −0.64 0.48 Iceland
0 0 −0.78 0.41 −2.88 0 Ireland
−0.81 0 −1.42 1.83 0 0 Italy
−1.37 0.99 −0.01 0 −1.98 2.17 New Zealand
−0.86 0.7 −1.11 1.94 −1.81 1.85 Norway
0 0 −2.88 0.89 −3.09 1.31 Phillippines
0 0 −1.47 1.88 −3.09 1.38 Poland
−0.15 0.14 −1.48 0.87 0 0 Saudi Arabia
0 0 −0.26 0 −2.33 2.75 Slovenia
−0.68 0 0 0 −2.23 0.01 South Africa
−0.04 0 0.27 1.74 −0.52 0.36 Spain
0.02 0 −1.02 0.5 −1.94 1.58 UK
−0.02 0 −0.73 −1.22 0 −1.49 US
Emo_F
Emo_M
nonver_F
nonver_M
obj_F
obj_M


Correlation between GGI & Permutation Test Z−score
nonver_F
obj_F
Emo_M
nonver_M
Emo_F
obj_M
Measurement
0.00
0.05
0.10
0.15
Correlation
positive
FALSE
TRUE

Discussion

Emotion expression difference between genders is not obvious for all of the 20 countries; nonverbal behavior differences between genders are more obvious, while objects in picture differences are significant cross most of the countries. differences in feminine nonverbal behaviors, feminine emotion expression and masculine objects are positively correlated with GGI; the difference in feminine nonverbal behaviors has the largest positive correlation with GGI. While differences in feminine objects, masculine nonverbal behaviors and masculine emotion expression are negatively correlated with GGI.

Conclusion

Gender differences in terms of composite variables seem consistent across all the countries; and GGI is obviously correlated with gender difference levels.

Ant Anatomy and Behavior Research

Introduction

The client of this project is Darcy Gordon, a PhD candidate from BU Department of Biology. In her experiment, there are 3 morphological groups of ants: minor(17 ants), major(18 ants) and supersoldier(17 ants). Each ant was tested with chemical trails of 3 concentrations: 0.0003, 0.001 and 0.003, so the whole data set has 3*(17+18+17) = 156 observations. Outcomes are binary: 1 means successful detection, 0 means failure. This project focuses on the effectiveness of chemical concentration in increasing detection rates; difference in detection rates of ants groups; whether concentration effectiveness changes with groups.

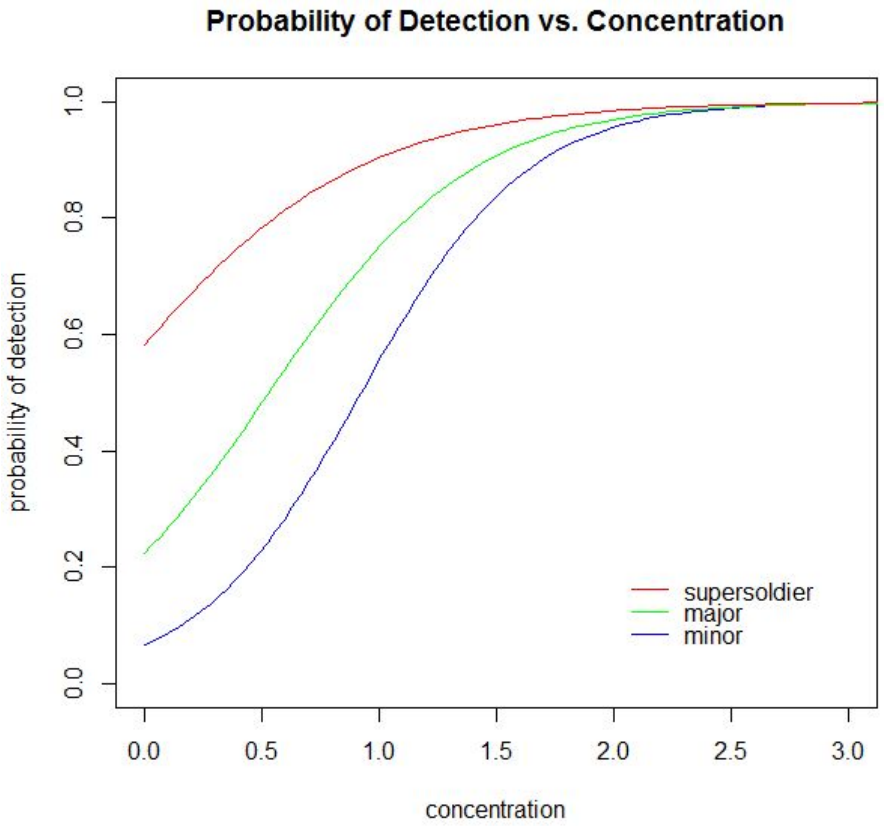
Materials and Methods

Raw data is in Excel format, main visualization and analysis is conducted in R. In the following equation, j indexes each ant, i indexes each trial; y indicates each experiment result: y is 1 if the ant successfully detect the chemical trail, y is 0 if it fails to detect. y_i follows a bernoulli distribution with success probability p_i. p_i could be explained by a linear combination of x_{is} and x_{ic}, here s indicates subcategory (categorical) and c indicates concentration (continues).

$$y_i \sim Bernoulli(p_i)$$
$$p_i = logit^{-1}(\beta_0 + \beta_s x_{is} + \beta_c x_{ic} + \beta_{sc} x_{is} : x_{ic} + \alpha_{j[i]})$$
$$\alpha \sim N(0, \sigma_\alpha^2), \text{ j = 1, 2, ..., J}$$

After fitting the model described above, expected probability of detection is plotted against concentration; results could be reached through visualization.

Results and Discussion



It is obvious that increase in concentration will result in increased detection rate. Ants from different groups have a distinction in detecting chemical trails: Supersoldiers have the highest detection rate and Minors have the lowest detection rate. What's more, we may tell that the effect of concentration is varied for different subcaste from the interaction term in our model, since the slopes of curves in the plot above change in different rates.

Conclusion

Concentration of chemical trails has influence over detection probability of ants, and the influence level varies for different groups.