Multinomial Models Trained via Mechanical Turk Surveys

By: Oseas Ayerdi and Wendy De La Rosa

Introduction

We conducted a survey via Amazon Mechanical Turk to gather data on salary and bank account balance levels for different demographics in the US; this information, in turn, was used to develop four multinomial models that predict said levels on several sectors of the US population. These results are then aggregated to provide a distribution for the US-level responses, which are presented and compared against figures published elsewhere.

Methodology

Survey

An eight-question survey was posted in Amazon Turk to better understand the salary levels, as well as use of savings and checking accounts for several sectors of the US. The first four questions asked for the respondent's gender, age, educational level, and race. The other four questions (*answer options*) were:

1. What is your monthly salary (if none, choose 0)? (*0/1-249/250-999/1000-2499/2500-4999/5000-9999/10000+*)
2. Which range below (in US Dollars, USD) matches the balance in your checking account (if no account, choose 0-9)? (*0-9/10-49/50-249/250-499/500-999/1000+*)
3. Do you have a savings account? (*yes/no*)
4. Which range below (in US Dollars, USD) matches the balance in your savings account (if no account, choose 0-9)? (*0-9/10-49/50-249/250-499/500-999/1000+*).

The respondents were compensated $0.04 cents for participating in the survey. We collected responses from 200 participants in a span of 11 hours. On average, each respondent took 32 seconds to complete the survey. Other than being a resident of the US, no other requirements were placed on the participants such as historical accuracy rate, educational level, etc; we wanted the survey to be opened to as-wide a range of residents in the US as possible.

Multinomial Models

The results from the survey were then used to train four multinomial models (`salary_model`, `checking_balance_model`, `savings_model`, and `savings_amount_model`) that predict the most likely response to the four questions above for all 256 combinations of age, education, sex, and race found in the `pop_dist.tsv` file. The predicted values along with the demographic information is saved in the file `hw4_predicted_values.csv`.

US-Level Responses

The stratified responses were then used to look at the distribution of the population of the US that would provide a specific response to the four questions discussed above. This was done in `R` via the `tapply` function. For example, for the salary level response, the proportion of the US population (`population_df$N`) that would give each specific response (`population_df$salary`) was added up, and then divided by the total US population present in the `pop_dist.tsv` file (`Total_US`). The sample code for this specific step is shown below:

```
US_salary = tapply(population_df$N, population_df$salary, sum) / Total_US
```

These results along with other more granulated observations are presented in the section below.

Results

US-Level Responses: Monthly-Salary Level

Level (in USD)	0	1-249	250-999	1,000 -2,499	2,500 -4,999	5,000 -9,999	10,000+
% of US adult population*	4.00%	8.23%	20.20%	42.17%	15.45%	9.65%	0.29%

*In the `pop_dist.tsv` file, numbers may not add up to 100% due to rounding.

The multinomial models predict that the majority of adults in the US have a salary of 1,000-2,499 USD, or roughly an annual salary in the range of 12k-30k USD; the results above also indicate that the median is somewhere in that range too. A report by the US Census Bureau (USCB) indicated that the median household income of the US in 2016 was roughly $59k USD[1]; in that same year, the USCB also concluded that the number of households in the US were roughly 125.82 MM[2]. If we use the total number of adults in our dataset (i.e. 248 MM), we get an estimate for the number of adults per household (i.e. 1.97), meaning that, per the USCB numbers, the average adult earns $30k USD, or in line with the calculation presented above.

US-Level Responses: Checking Account Balance

Level (in USD)	0-9	10-49	50-249	250-499	500-9999	1000+
% of US adult population*	7.13%	14.37%	26.79%	9.66%	8.30%	33.75%

*In the `pop_dist.tsv` file, numbers may not add up to 100% due to rounding.

Our results predict that the majority of adults in the US have a balance in the range of 50-249 USD. A comparable figure was reported by ValuePenguin[3], in which they provide a mean figure for checking account balance in the US; their number, however, is considerably higher: 2,900 USD. This can be a result of bias in our sampling, as one can imagine that individuals participating in alternative means of income such as Mechanical Turk may have less disposable income than those that do not.

US-Level Responses: Savings Account

Response	no	yes
% of US adult population*	18.58%	81.42%

*In the `pop_dist.tsv` file, numbers may not add up to 100% due to rounding.

Our results indicate that roughly 19% of the US population (aged 18+) —or a bit over 46 MM people—do not have a savings account. A proxy for a comparable figure can be obtained by using two reported values. One of these is the % of respondents that reported a balance of 0 in a GoBankingRates[4] survey for the balance in their savings account; this figure is 39%. The second figure is the portion of adults that

[1] https://www.census.gov/newsroom/press-releases/2017/income-povery.html
[2] https://www.statista.com/statistics/183635/number-of-households-in-the-us/
[3] https://www.valuepenguin.com/banking/average-checking-account-balance
[4] https://www.gobankingrates.com/saving-money/half-americans-less-savings-2017/

have banking services, which can be obtained by removing the estimated FDIC[5] figure for the number of adults in the US without any banking services (15.6 MM) from the total population in the `pop_dist.tsv` dataset. This gives the comparable figure of 90 MM (i.e. $38\% * (Total_{us} - 15.6)$) people, which is twice as large as that from the model. Like the results in the salary level section, this can be due to bias in the survey demographics, with people already having a checking account as a means to get compensated via Mechanical Turk, are more likely to also have a savings account than those that have neither.

US-Level Responses: Savings Account Balance

Level (in USD)	0-9	10-49	50-249	250-499	500-9999	1000+
% of US adult population*	49.66%	21.32%	0.77%	0%	3.41%	24.84%

*In the `pop_dist.tsv` file, numbers may not add up to 100% due to rounding.

We compare our results above with a 2017 GOBankingRates[6] survey replied to by over 8,000 Americans. One of the findings in said survey was that over 57% (and 69% in their 2016 survey) of the respondents had a savings account balance of less than 1k USD; the number from our results is roughly 75.16%. A second observation from the GOBankingRates survey was the percent or respondents with a balance of 0% in their account (39%); our closest comparable number (% of those answering 0-9) is nearly 50%. Though these numbers are not exactly near each other, they are in the right order of magnitude, i.e. both conclude that the majority of adults in the US (>50%) have a savings account balance lower than 1k USD, and a significant portion of those have virtually no savings!

Limitations

There are several limitations in this report. To begin with, there is no way to verify the validity of the survey responses. The fact that the majority of these surveys were filled at a pace of 32 seconds per is a sign that the data can be less-than-reliable. In addition, there seems to be some selection bias in the people participating in the survey, given that the compensation of $0.04 cents translates to roughly an hourly rate of $2.40. In addition, our analysis with the exception of the salary compares our results with results from entities that may not have the resources and the scope that the figures from the US Census Bureau have, though more importantly than focusing on whether the numbers match exactly, it is more important to see if they tell a comparable story. With that being said, in the aggregate there were some comparable results with other published figures, in particular with the findings for the salary levels in the US.

[5] https://www.fdic.gov/householdsurvey/2015/2015execsumm.pdf
[6] https://www.gobankingrates.com/saving-money/half-americans-less-savings-2017/