Social Salary Study

S. Arora, J. Harmse, V. Mulholland

April 22, 2018

Study Overview

The purpose of this study to measure whether a person is driven by money or not. We found it reasonable to assume that a person who is driven by money would expect to earn more than the average person who has the same skillset and experience. The question to be answered is the following:

Is a person's social standards correlated with a person's drive for financial wealth?

The null hypothesis is that the people who are driven by financial wealth and those that are driven by job satisfaction share the same social standards in terms of spending patterns while the alternative is that there is enough evidence to support that the two groups do not share the same spending patterns. A logistic regression is used to model the probability of belonging to the said groups given the explanatory spending and country of origin variables.

Our survey captures the salary of what a participant thinks an average person with their skills and experience should earn, as well as the salary that the participant expects to receive in 1 year's time. The average inflation rate world-wide was of approximately 3.15 per the Statista Statistics Portal. Taking inflation and other micro-factors into account, a participant's expected salary in a year's time shouldn't be much higher than the average person with the same skills and experience.

The survey was answered by people from various countries with different currencies and it records the participant's salary in their respective currency. This means that we cannot compare the captured salary values between participants. An easy way of standardising these values is to handle the salary values as a ratio of the difference between expected salary and average salary over the average salary. The ratio should be consistent across different currencies.

For this study, social standards is defined as a person's inclination for a high relative consumption on leisure activities and non-essential expenditure. Our hypothesis relies on the theory that prevailing social conditions will influences one's relationship with money which would translate in whether increase in income is the priority. The study is inspired by Hirsh's hypothesis which states that prevailing social conditions will influences one's spending on status (status-seeking) goods and activities which would translate in those with a higher disposable income being more inclined to spend on non-essential expenditures.

Methodology

Survey Study Design

A survey was conducted on *Survey Monkey* to address the question at hand. The intended audience was originally our MDS class, but the invitation was extended to our social media accounts (LinkedIn, Facebook, etc.). The 5 questions in the Appendix under Survey Questions are conceptualized from both pertinent topics, one pertaining to the salary motivations and the latter is a measure of a participants social standards. This is admittedly a difficult concept to measure, thus our focus is mainly to delineated the difference between essential expenditure versus lifestyle enhacement spendings. The vacation spending category is considered a non-essential category since it can be extrapolated that any non-conspicuous travels would (such as traveling to a nearby town to visit relatives) will only make up a marginal proportion of the overall spending. The paper Conspicuous Consumption Applied to Tourism Destination states that the location or purpose of vacation will help determine whether the spending is categorized as conspicuous or not. The daily leisure,

hobbies and consumer goods catgories are all deemed to fall under the non-necessary spending categories and the living expenses, savings and other are classified as basic needs. The Study Design discussion will elaborate on whether these assumptions were truly justified at the phase of study design.

Data Collection Methods

Survey Outliers talk about how answer inputs were chosen.

Certain safety checks are put in place to prevent users from entering invalid data. For instance, the spending categories have to add up to 100% of their expenditure

The following table summarises the key fields populated by the survey data and the calculated value, `ratio`, namely the response variable as a ratio of the two salary values.

Features	Description
`salary_base`	An indicator meant to be a subjective baseline of what salary a person of their expertise would earn.
`salary_expect`	The expected salary combined with the base salary provides a relative indicator to the respondents pursuit of monetary gains.
`no_increase_acceptance`	A binary metric used to flag those that are more driven by money than others.
`ratio`	$\frac{Salary_{expect} - Salary_{base}}{Salary_{base}}$ allows us to differentiate those who have a desire for a high increase in salary vs those that are satisfied with a modest amount.
`living_expenses`	Living Expenses (utilities, rent, mortgage, transportation, property taxes if owner, etc.) percentage of spending
`savings`	Savings (retirement, investments, emergency funds, etc.) percentage of spending
`vacation`	Vacation (lodging, transportation, day trips, etc.) percentage of spending
`daily_leisure`	Daily Leisure (eating out, books, movies, self-care, etc.) percentage of spending
`consumption_goods`	Consumption Goods (clothing, electronics, other luxury items, etc.) percentage of spending
`sports_hobbies`	Personal Sports and Hobbies (sporting goods and services, gym, arts and crafts, etc.) percentage of spending
`other`	Other (health care, taxes, dependent expenses, etc.) percentage of spending

The data is anonymized before being made visible and then uploaded as a processed dataset. Several pre-processing steps are performed to clean and wrangle the data into a usable format. Our response variable underwent a transformation to better differentiate the participants of interest since we cannot assume that a person with a high salary ratio is primarily driven by wealth as they may have specified that job satisfaction is a bigger driver than salary increase and vice-versa. Those that do not fall into either these two groups are considered to have answered ambiguously (expects a high salary increase, but is content with job satisfaction or the opposite) and cannot be considered for our study.

$$\begin{cases} Wealth & \text{for } no\_increase\_acceptance = No \text{ and } ratio > median(ratio) \\ Not\ Wealth & \text{for } no\_increase\_acceptance = Yes \text{ and } ratio <= median(ratio) \\ Other & \text{otherwise} \end{cases}$$

Analysis Methods

From the exploratory analysis, it is determined that the response variable should be modeled using a logistic regression that identifies the participants that belong to one of two groups. The first group is comprised of respondents that are deemed driven by financial wealth by indicating that they would not prioritize job satisfaction over a raise within two years and their salary ratio is above the median. The second group is at the opposite end of the spectrum specifying that they would maintain a job given satisfaction and that they expect less than the median salary ratio.

Before settling on the model features, the confounding variables had to be assessed. From the EDA performed in the Appendix Continent section, it seems viable that a person's origin may influence their spending habits and relationship to money out of cultural differences, which we would want to account for.

Figure **??**.

INSERT MODEL HERE

results and analysis

The Anova test results in the Appendix Continent section has the `Pr(>Chi)` output indicates 0.05046. Since this value is on the verge of being below the significance level and the fact that we do not have a very large number of observations, the `continent` variable will be used in our model as a confounding variable.

Discussion

Study Design

The Survey Questions were constructed to account for all types of spending, so that the respondent could better consider their proportional spending distribution. There are many subjective and psychological features that would contribute to someone's self-assessment of expected and base salary estimates which was accounted for when stating that we are looking at a person's drive for money, therefore their perceived attitude toward financial vocational incentives.

Clarifying the spending categories is a shortcoming of our study. There is a tradeoff between making our survey straightforward and being too transparent about the agenda behind the analysis with overly specific questions. There is some ambiguity behind the concept of social standards which we tried to account for in the vacation, hobbies, consumption and daily leisure categories, but we acknowledge that one could be partaking in conspicuous consumption while categorizing it as a living expense, such as paying a very high rent to live in the nicest neighbourhood. The "Other" category could also be misleading because there could be some frivolous expenses that are not accounted for.

Self-assessments aren't ideal since the participant is require to think objectively on the spot about their finances. This could inject a considerable source of bias, and would have required a more thorough assessment method than a survey. Providing an export of a bank categorization of one's spendings would be a better method for a true representation.

It was good to use the point system to divide the spending because it forced the participant to consider each category of interest in proportion to the others. The improvement we would make is engineering a clearer divide between what is considered basic needs or not and comparing the expenditures to the local average spending percentages on necessities versus conspicuous spending.

Results Discussion

Conclusion

Appendix

Survey Questions

1. What is your country of employment/future employment? (used for determining currency for following questions)
2. Assuming the country's currency specified above, what should someone with your qualifications and experience expect to receive as an annual salary?
3. Assuming the country's currency specified above, what is the annual salary that you aim to receive 1 year from now?
4. Assuming high job satisfaction, would you keep a job that does not give you a salary increase over the next two years?
5. Please assign an approximate percentage of your current yearly expenses to the following categories (must sum up to 100).
    - Living Expenses (utilities, rent, mortgage, transportation, property taxes if owner, etc.)
    - Savings (retirement, investments, emergency funds, etc.)
    - Vacation (lodging, transportation, day trips, etc.)
    - Daily Leisure (eating out, books, movies, self-care, etc.)
    - Consumption Goods (clothing, electronics, other luxury items, etc.)
    - Personal Sports and Hobbies (sporting goods and services, gym, arts and crafts, etc.)
    - Other (health care, taxes, dependent expenses, etc.)

EDA

Confounding Variables

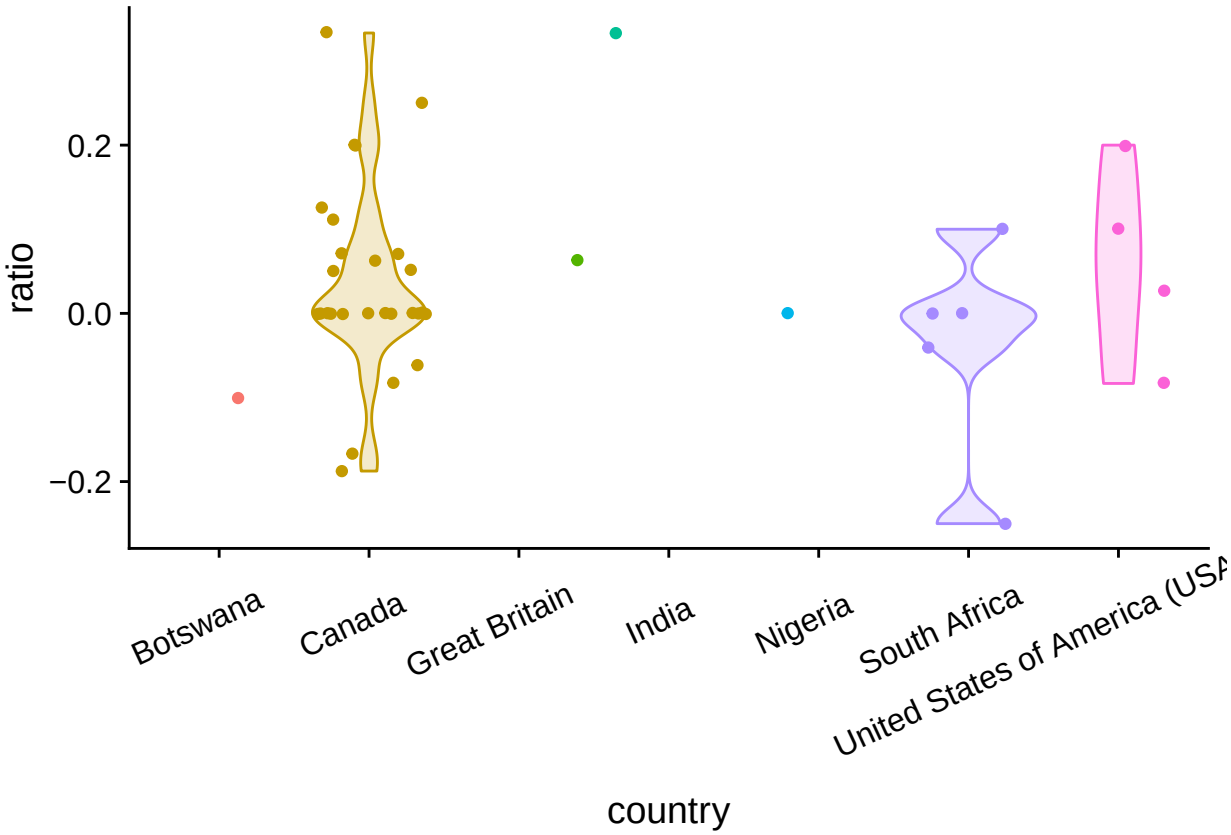
Continents

We want to determine whether people who are driven by wealth have different spending habits than people who are not driven by wealth. The nature of the hypotheiss and the data makes a logistic regression model the obvious choice for determining whehter any expense category is significantly different between the two groups.

Before we start building the model we should consider whether we are dealing with any confounding variables.

Each person who completed the survey had to report their country of residence. Data was collected from a number of different countries. The study's response variable is standardised salary ratio, which does not require taking the person's country into account. However, the country a person live has the potential to play a role in a person's spending habits regardless of their salary ratio. For example, a person from Africa would not necessarily spend a lot on vacations in comparison to a person from North America which may be a result of something like cultural differences.

We need to consider whether country has any significant effect on either out explanatory expense variables or our response variable.




```
## Analysis of Deviance Table
##
## Model 1: group_bin ~ living_expenses + savings + vacation + daily_leisure +
##     consumption_goods + sports_hobbies + other
## Model 2: group_bin ~ living_expenses + savings + vacation + daily_leisure +
##     consumption_goods + sports_hobbies + other + continent
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        31     38.156
## 2        30     34.329  1   3.8262  0.05046 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```