Subjective Well-Being Data Task

This document is meant to be self-contained, which means that there should be no need to use outside material for any of the questions presented below. You may use any programming language (Stata, R, Python, etc.). Please submit your code script along with your answers to the questions in a separate document (PDF or docx).

Background

Inequality has been a growing topic of discussion for economists and for society more broadly. Income inequality has received the bulk of the attention, but we may also care about inequality as is pertains to non-market goods such as "how rewarding your life is" or "your sense of security". For this data task, you will work with data collected from a survey designed to study exactly this issue of measuring interpersonal differences for a broad class of non-market consumption goods which we will call aspects of well-being.

In this survey, respondents answer one main type of question, which we will call a rating question. Figure 1 displays an example of what a respondent sees when answering these questions.

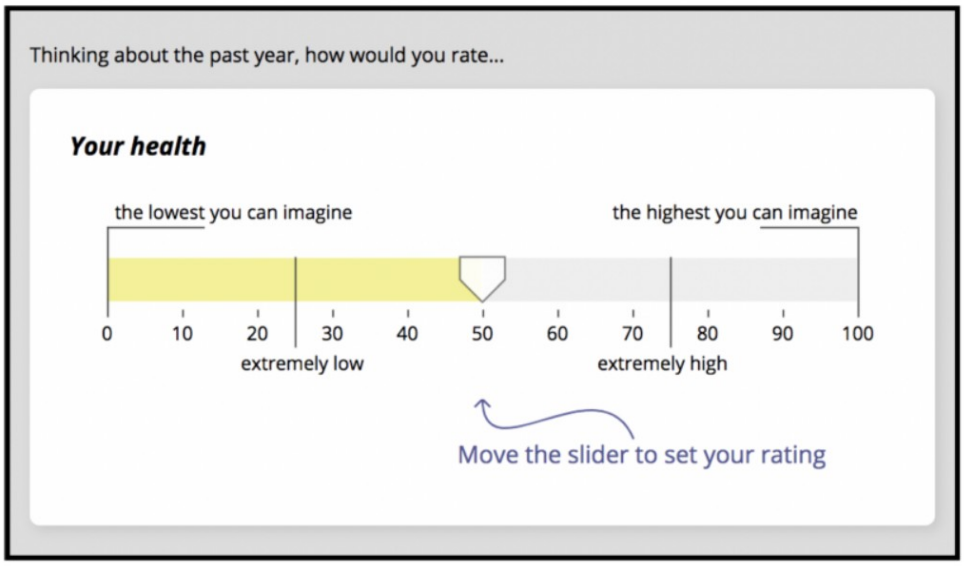

Figure 1: Example Rating Question

Note that respondents rate aspects on an integer scale from 0 to 100. Respondents are given no direction on how to answer these questions other than a short set of instructions copied here:

> Please rate the aspect on a scale from 0 to 100, where 0 is the lowest level you can imagine in anyone's life, and 100 is the highest level you can imagine in anyone's life.
>
> Note: All the aspects in this survey are worded so that for most people, a higher rating is better and a lower rating is worse.
>
> Also note: A rating of 25 is extremely low and 75 is extremely high, so a rating above 75 or below 25 is beyond extreme.

In this data set, you have ratings from respondents who each rate the same set of aspects. The aspects are designed to cover a broad range of what people care about in life. In addition to rating questions, respondents also report their demographic information (gender, age, etc.).

Data

You have been provided two data files: `ratings.csv` and `demographics.csv`. The column names and variable descriptions for both files are listed below:

Column name	Description
`aspect`	Name of aspect of well-being
`worker`	Unique string of letters and numbers identifying respondent
`time`	Time that the respondent recorded his/her rating. Reported in Unix time (Number of seconds since Jan. 1 1970) more info
`rating`	Rating on a scale from 0 to 100
`age`	Age of respondent
`male`	Binary variable for respondent gender. 0 indicates female and 1 indicates male
`income`	Respondent's total household income
`education`	Respondent's completed level of education
`race`	Race of respondent

Question 1

For this question, you will work with the ratings data set. Please conduct the following steps

(a) Load `ratings.csv`

(b) Report the number of unique respondents and the number of unique aspects in the data set

(c) Check to see if each respondent has only rated each aspect once. If this is not true, only include the most recent observation and report the number of observations you have dropped.

(d) Calculate the average rating for each respondent. We will call this measure subjective riches. Report the minimum, 25th percentile, 50th percentile, 75th percentile, and maximum subjective riches value.

Question 2

(a) Load `demographics.csv`

(b) Report the number of rows and check to see if it is the same as the number of unique respondents you calculated in question 1.

(c) Merge the subjective riches data from question 1 with the demographics data.

(d) Regress (with OLS) subjective riches on income and report the results.

- Interpret the results. What is the relationship between income and subjective riches? (Max 100 words)

(e) Regress (with OLS) subjective riches on income with controls for age, age^2 (age squared), gender, level of education, and race.

- Interpret the results. (Max 150 words)

(f) Imagine you were also given each respondent's household size. How would you change your analysis above in light of this new information? (Max 100 words)

Question 3

Your PI is giving a presentation to a health-policy audience, and she would like to display a figure that illustrates the relationship between subjective ratings of health, income, and age. She has asked you to produce a single scatterplot that conveys the relationship between all three variables.

(a) List the steps you would take to produce the scatterplot. Remember

- Any individual/average rating data in your plot should be for aspects related to health.
- All three variables should be featured in some way on the plot.

- The figure should readable, effectively convey the information through visuals, and preferably be intuitively understandable to an audience that has limited familiarity with the survey and your data set.

(b) Produce and save the scatterplot (or if you prefer, up to two proposals for alternative scatterplots).

(c) From a policy perspective, understanding the determinants of well-being is an important question. Describe the ways in which your regressions in the previous question and your scatterplot(s) help or do not help answer this question. Think about your proxy for well-being (subjective ratings) as well as the specification of your regressions. (Max 250 words)