Subjective Well-Being Data Task

Introduction

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, MATLAB, etc.). Please submit your code script along with y our answers to the questions. If you come across any problems or errors, you can email dleksanov@uchicago.edu. Note that any word limits on answers will be enforced by truncating your response at the point that the word limit is crossed.

Please do not include your name, email, or any identifying information in your submission.

We grade data tasks blind to make sure these grades aren't influenced by other parts of your application.

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 it 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.

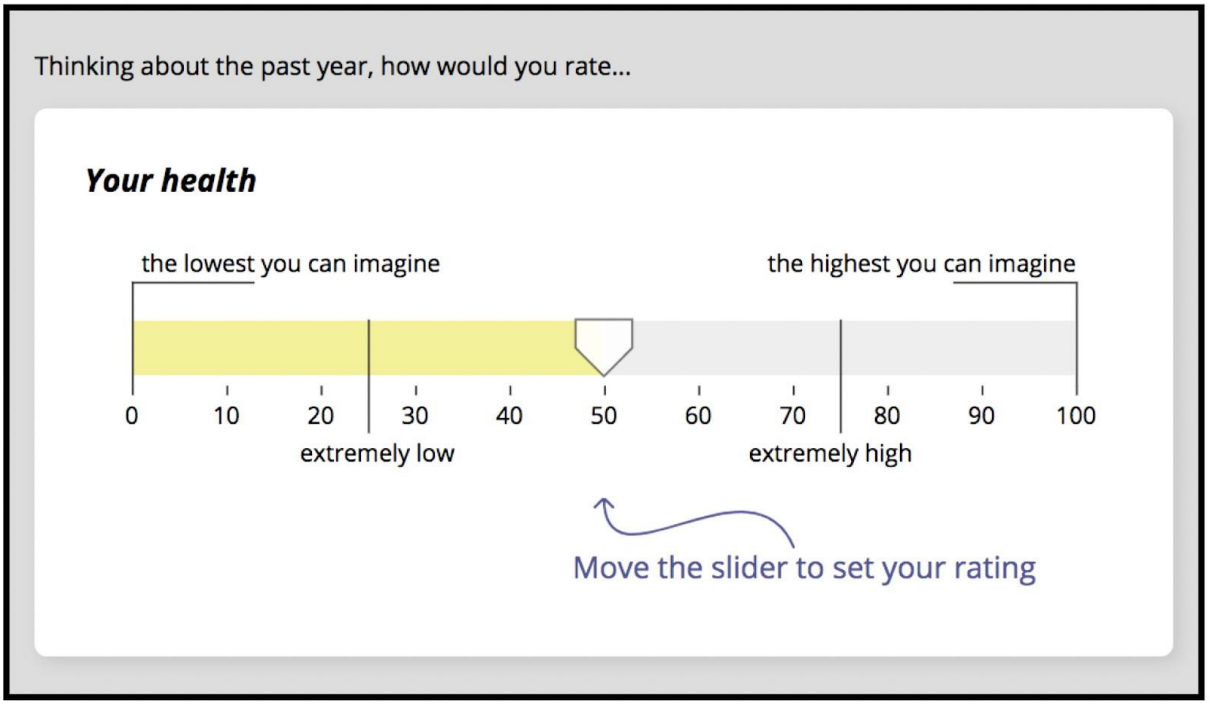

Figure 1: Example Rating Question

Note that respondents r ate 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.
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 dataset. Please conduct the following steps:

(a) Load *ratings.csv.* Report the number of unique respondents and the number of unique aspects in the dataset. [7 points]

(b) 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 (i) the number of observations you have dropped and (ii) the average time (no conversion needed) of all observations left. [7 points]

(c) 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. [6 points]

Question 2

(a) Load *demographics.csv.* Report the number of rows. Is this the same as the number of unique respondents you calculated in question 1? [3 points]

(b) Merge the subjective riches data from question 1 with the demographics data. Report the number of unique workers and the total number of rows in the new dataset. [6 points]
(c) Regress (with OLS) subjective riches on income and report the results. [10 points]
    (i) Interpret the results. What is the relationship between income and subjective riches? (Max 100 words)
(d) Regress (with OLS) subjective riches on income with controls for age, age ^{2}, gender, level of education and race. [11 points]
    (i) Interpret the results. If the results in this regression differ from the results in 2d, talk about possible reasons for this and support your suggestions with data. (Max 150 words)
(e) 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) [10 points]

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 scatter plot that conveys the relationship between all three variables.

(a) List the steps you would take to produce the scatterplot [10 points]. Remember:
    (i) Any individual/average rating data in your plot should be for aspects related to health.
    (ii) All three variables should be featured in some way on the plot.
    (iii) You may include trend-lines as part of the scatterplot.
    (iv) The figure should be 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. [10 points]
(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 to answer this question. Think about your proxy for well-being (subjective ratings) as well as the specification of your regressions. (Max 250 words) [20 points]