Subjective Well-Being Data Task

Jason Cruz
164468@unsaac.edu.pe

Setiembre, 2021

Question 1

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.

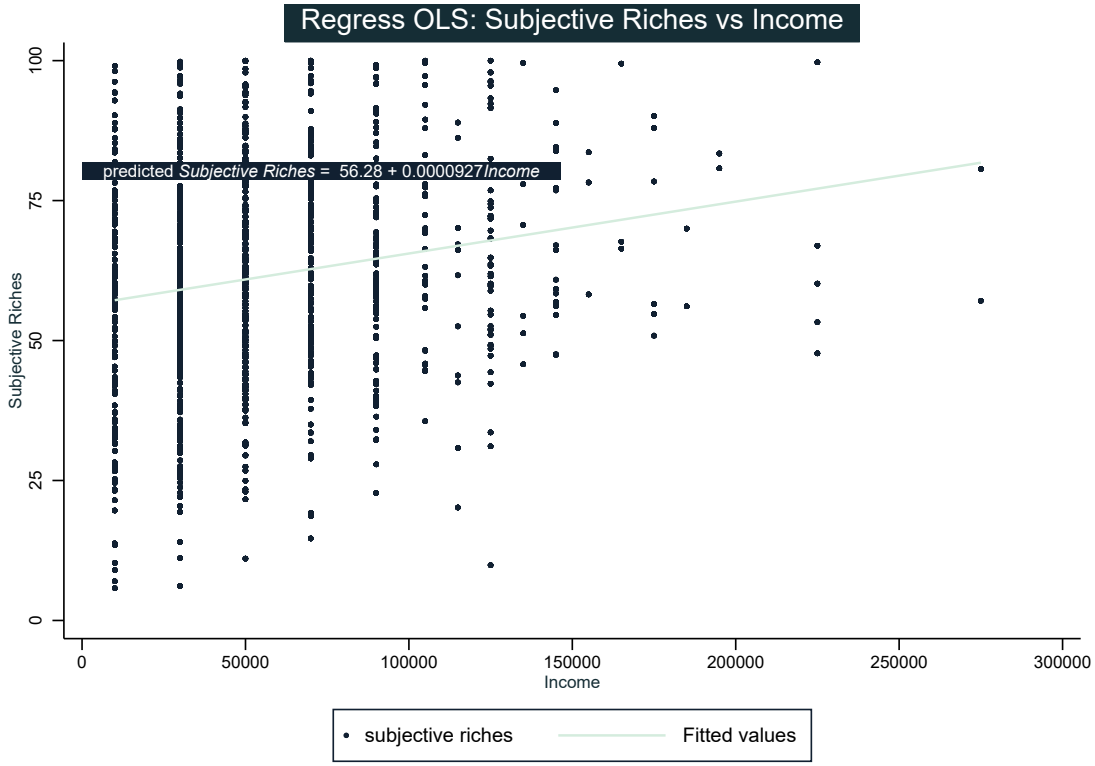
Question 2

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.



	(1)	
	Coef./S.E.	p-value
income	0.00***	0.00
	0.00	
Const	56.28***	0.00
	0.25	
R-cuadrado	0.04	
N. de obs.	17952.00	

* $p < 0{,}05$, ** $p < 0{,}01$, *** $p < 0{,}001$

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

  There is a very weak positive relationship. The coefficient 0.0000927 (significant by p-value) is interpreted as: for each additional monetary unit that the respondent receives, his or her score in aspects of well-being increases by 0.0000927 points. This low value means that there are other determinants other than income that better explain the differences in aspects of well-being (such as health, happiness, etc.). In addition, being a bivariate model without controls, the correlation is weak with a very low R squared, although this is not a dazzling indicator in this analysis, we are practically facing an ad hoc model.

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

	(1)		(2)	
	Coef./S.E.	p-value	Coef./S.E.	p-value
income	0.00***	0.00	0.00***	0.00
	0.00		0.00	
Const	56.28***	0.00	68.66***	0.00
	0.25		1.89	
age			-0.35***	0.00
			0.07	
age_squared			0.00***	0.00
			0.00	
0.male			0.00	.
			.	
1.male			2.57***	0.00
			0.29	
1.education_new			0.00	.
			.	
2.education_new			-6.90***	0.00
			1.33	
3.education_new			-7.61***	0.00
			1.27	
4.education_new			-5.68***	0.00
			1.49	
5.education_new			-4.55***	0.00
			1.27	
6.education_new			-7.28***	0.00
			1.35	
7.education_new			-2.39	0.12
			1.54	
1.race_new			0.00	.
			.	
2.race_new			7.11***	0.00
			0.70	
3.race_new			-2.33***	0.00
			0.55	
4.race_new			-2.02***	0.00
			0.56	
5.race_new			-0.67	0.21
			0.54	
6.race_new			8.41***	0.00
			1.62	
R-cuadrado	0.04		0.06	
N. de obs.	17952.00		17935.00	

* $p < 0{,}05$, ** $p < 0{,}01$, *** $p < 0{,}001$

- Interpret the results. (Max 150 words).

The model has improved since the specification, it is more interesting the relationships of the control variables with subjective riches according to the results. For example, the negative coefficient of age (significant) shows that each additional year represents -0.35 points on the aspects of well-being, while the coefficient of age_squared shows that at a certain age (many years of life) the respondents value more the aspects of well-being, older adults probably rate well-being aspects higher than young people. Likewise, being a male respondent means scoring the aspects of well-being with 2.57 points more than women. On the other hand, for the most part, the categories of the variables education and race score the aspects of well-being less.

f) Imagine you were also given each respondent's household size.(Max 100 words).

	(1)		(2)		(3)	
	Coef./S.E.	p-value	Coef./S.E.	p-value	Coef./S.E.	p-value
income	0.00***	0.00	0.00***	0.00		
	0.00		0.00			
Const	56.28***	0.00	68.66***	0.00	7843.48*	0.04
	0.25		1.89		3830.02	
age			-0.35***	0.00	243.69	0.10
			0.07		149.33	
age_squared			0.00***	0.00	-4.46**	0.01
			0.00		1.67	
0.male			0.00	.	0.00	.
			.		.	
1.male			2.57***	0.00	616.53	0.29
			0.29		583.87	
1.education_new			0.00	.	0.00	.
			.		.	
2.education_new			-6.90***	0.00	9084.86***	0.00
			1.33		2203.85	
3.education_new			-7.61***	0.00	17652.57***	0.00
			1.27		2141.13	
4.education_new			-5.68***	0.00	49033.29***	0.00
			1.49		2580.31	
5.education_new			-4.55***	0.00	29767.44***	0.00
			1.27		2141.46	
6.education_new			-7.28***	0.00	49755.59***	0.00
			1.35		2410.27	
7.education_new			-2.39	0.12	54239.02***	0.00
			1.54		3910.63	
1.race_new			0.00	.	0.00	.
			.		.	
2.race_new			7.11***	0.00	-2116.96	0.18
			0.70		1586.42	
3.race_new			-2.33***	0.00	-2897.10*	0.01
			0.55		1130.09	
4.race_new			-2.02***	0.00	-5405.79***	0.00
			0.56		812.69	
5.race_new			-0.67	0.21	12723.95***	0.00
			0.54		1428.59	
6.race_new			8.41***	0.00	-24345.41***	0.00
			1.62		2995.12	
subjective_riches					350.85***	0.00
					13.95	
household_size					165.15	0.40
					197.27	
R-cuadrado	0.04		0.06		0.14	
N. de obs.	17952.00		6 17935.00		17935.00	

* $p < 0{,}05$, ** $p < 0{,}01$, *** $p < 0{,}001$

How would you change your analysis above in light of this new information?

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)