

Day 4 Group exercise: Survey research

DATA5207: Data Analysis in the Social Sciences

Dr Shaun Ratcliff

In both labs today, we will be working on an assessable group exercise. You will have a day to finish and upload your work (see the *Group assignment 1* assessment on canvas).

This task, which counts towards your group work grade, should take you no more than three or four hours to complete (most of which will be time we give you to work on this in the labs). It is worth 10 per cent of your total grade for the unit and will be conducted in your groups for the assessable group projects, which should have approximately four or five members. You should submit your final written document as a *Markdown* file through a link that will be provided in this module. You do not need to knit this, just submit the RMD. Only one member of your group needs to submit your work. As long as the groups are properly registered on canvas, all members of the group will receive the full grades awarded.

The assessable group task for today

You will conduct an analysis on life quality using the *World Values Survey*.

Your goal in this project is to come up with a theory that explains quality of life. You are required to work out what might best predict these quality of life scores.

To do this we will give you a survey and corresponding data (The World Values Survey). You will use these data to identify predictors for quality of life. **We will create the dependent variable in the second lab.** This will be a measure of quality of life created with a combination of four questions from the survey using factor analysis. **This morning, you will develop theories for what might predict higher/lower quality of life.** Think about what drives life quality. Look at questionnaire and codebook. Identify possible predictors. Write up the descriptive analysis. Means, sd, or histograms, tables.

This afternoon we will create your dependent variable and you will test your ideas. Further details are available below.

Part 1: Understanding your survey data

This morning, you will be familiarising yourself with the survey data, and examining possible predictors, or independent variables, for this project.

You will undertake three main steps:

1. Develop a theory or theories of what factors might influence quality of life, and why. This will inform your selection of independent variables (predictors). This does not need to be sophisticated or complicated. A few simple ideas will suffice. Write this up in the Markdown file you will be submitting.
2. Using this theory, select what you believe to be appropriate predictors to explain quality of life from the provided survey data. There should be at least five of these.
3. Examine these predictors in *R*. Look at their mean, standard deviation, distribution, etc. Do they need to be recoded in any way? Write up your data cleaning decisions in the Markdown file, and include any descriptive analysis you do.

Each step in your Markdown file should be clearly labelled.

Part 2: Using factor analysis

We will begin the afternoon lab by creating a measure for life quality from five variables in the the World Values Survey. This will be done with factor analysis, and we will walk you through this process during the first half of the lab. In the second half of the lab, you will use the variables you selected this morning to predict respondents quality of life. How well do your theories hold up?

For this section of the project, you will also undertake three main steps:

1. Calculate and plot the relationship between your dependent variable (the quality of life measure we calculated above using factor analysis) and independent variables using descriptive statistics.
2. Run linear regressions on the quality of life measure. Use appropriate measures to test your theories.
3. Write up your key findings.

Submit your work

Once you are finished, submit the RMD you are working on using the *Group assignment 1* assessment on canvas, along with your data. Only one version needs to be submitted. Full grades will be allocated to all registered members of the group.

All the files needed to run your code should be uploaded in a zipped folder. We should be able to run the code without changing it.

You will be marked on the quality of your *R* code (including whether it runs for us without errors), how well you have justified your variable selection, and the proper use of appropriate methods (see the rubric available on the Assessment information page).

If you need help

If you have any questions, do not hesitate to ask us for help. During the labs, the teaching team will be available to talk you through the project on Zoom. We will be moving through the breakout groups and can also be called to assist you. Outside of the sessions today, you can also post questions on the Ed discussion board.

We cannot do the work for you — this is an assessment – but we provide some advice.

Good luck with the exercise!