

Instructions for ST201 Group Project (worth 20 marks): Please read these instructions carefully.

Yunxiao Chen (y.chen186@lse.ac.uk)

Please post any queries on the forum. I will make it my priority to answer questions about the coursework. I will not answer any emails directly unless they are personal (e.g., you are unwell and cannot complete the work).

Background

The COVID-19 pandemic is associated with threats to the health and life of people around the world. It also leads to changes in everyday functioning, many of which are perceived as negative. In this project, the aim is to understand the relationship between people's generalised anxiety and their age, sociodemographic variables (e.g., gender, education, marital status), perceived health and life risks of COVID-19, pandemic-related difficulties and social support.

The goal is to understand the following questions:

1. Given the other predictors, is there a significant association between people's generalised anxiety and their age?
2. What is the relationship between perceived health and life risks of COVID-19 and people's mental health status? Does this relationship vary across different age groups?
3. Establish a linear regression model to summarise and interpret the relationship between generalised anxiety and the available predictors.

Data

The data include a total of 1115 participants (aged 18–85) who took part in a social study. The sample was representative of the population of a European country in terms of sex, age, and place of residence. The following variables are available in the dataset.

Demographic variables:

1. Sex (1= female and 0 = male)
2. Age

3. Education (1 = Primary education 2 = Vocational education 3 = Secondary education 4 = Post-secondary education 5 = University degree and above)
4. IncomeContinuity (whether they had income continuity during the COVID-19 pandemic 0 = no, 1 = yes)
5. HealthStatus (whether they suffered from any pre-existing medical conditions (chronic diseases, immuno- suppressive therapy, etc.) that may put them at risk of severe COVID- 19 infection. 1 = yes, 2 = don't know, and 3 = no.)
6. Unemployed (Whether they are unemployed 0 = no, 1 = yes)
7. Student (student status 0 = no, 1 = yes)

Pandemic difficulties:

To what extent each of the following is currently a problem for you? (1 = it is not a problem at all to 4 = it is definitely a problem for me)

Household relationship difficulties

8. Pandemic_Difficulties_1: Difficult relationships at home (feeling that we “act on our nerves”, quarrels, etc.)
9. Pandemic_Difficulties_2: Inability to stay alone (without other people)
10. Pandemic_Difficulties_3: Increased number of daily duties
11. Pandemic_Difficulties_4: Loneliness, a feeling of being abandoned

External restrictions

12. Pandemic_Difficulties_5: Feeling that my freedom as a human is restricted (e.g. deciding for myself)
13. Pandemic_Difficulties_6: Change in the way of shopping due to new regulations (e.g. reduced number of people in a store, the need to shop online)
14. Pandemic_Difficulties_7: Wearing face masks in public places

Lack of social contacts

15. Pandemic_Difficulties_8: No meetings with beloved ones and friends
16. Pandemic_Difficulties_9: No meetings with people in general

Fear and uncertainty

17. Pandemic_Difficulties_10: Feeling of uncertainty, unpredictability of the situation
18. Pandemic_Difficulties_11: Feeling of danger, anxiety associated with the spread of the virus

Others

19. Pandemic_Difficulties_12: Restriction in moving around and traveling
20. Pandemic_Difficulties_13: Boredom, monotony
21. Pandemic_Difficulties_14: The need to change my lifestyle
22. Pandemic_Difficulties_15: Feeling tired of this whole situation
23. Pandemic_Difficulties_16: Feeling lost in the recommendations and restrictions (e.g. getting lost in what you can and cannot do)

24. Covid_risk: Covid-related risks -- the subjective (i.e., perceived) risk (1 - very low to 5 - very high) of COVID-19 infection and its adverse health effect
25. Social_support: Social support – the amount of social support the participants received during the pandemic (1 - very low to 5 - very high)
26. Gad_score: Generalised anxiety level -- a standardised continuous variable where a large value indicates a high anxiety level.

Coursework Submission

Please attach a cover page when submitting your report. You can find a template for the cover page on Moodle.

Group work and log keeping (1 marks)

You have already formed groups. Please work with your group members. Changing groups is not allowed, except for exceptional circumstances (decided case by case). I provide a template on Moodle so that you can log your group meetings, who attended, what was discussed and what the goals are. It is very brief. You will hand in this log together with your project – you could be docked marks if you do not. This will ensure a record of who attended the meeting (this can be done remotely if the group is happy with it), who was allocated what works, and whether they did it. The log must be agreed upon by all present group members and referred to at the beginning of each meeting. This is similar to the concept of minutes which are records kept in meetings in all places of employment.

Report (19 marks)

This report will need to describe the regression analysis you conduct to answer the above research questions and other sensible research questions you can raise. The total word count is 2500, excluding the tables, figures, and appendix.

Here are several suggestions from me regarding the writing:

- Be factual – do not invent things! Stick to what you have learned from the course.
- Be clear – use simple and clear language. It is hard for me to assess your understanding if you use complex language and exaggerations.
- Have a smooth flow – make sure your report reads smoothly. Do not use notation, acronyms, terminology that have not been defined before. Do not suddenly jump to a topic without connecting with the previous discussion.

Do Not

- Include R code output in the report. Tables should be used to display results.
- Go over the stated word limit. Choose what you say carefully.
- Write irrelevant stuff to increase the word count until you hit the limit. Concise and clear reports are best, and you are conveying factual information, not writing a novel.

The report should contain the following sections:

1. Introduction (2 marks)

The introduction should be a summary of the research. First, it should clearly state the questions we are trying to answer and why we are interested in these questions. Second, you should give a brief background on the data and why it is appropriate to use it to answer the questions. Finally, briefly describe the methods you use in this research and explain why they are suitable here.

2. Data and Exploratory Data Analysis (2 marks)

Provide more details about the variables in the dataset and conduct an exploratory analysis of the data. For example, you can examine the data missingness at this stage and decide how to treat the missing data. You may also want to draw plots to look at the relationship between each predictor and the response variable, suggesting transformations or functional forms to be used in the regression model. You should only present the plots that drive certain decisions in your subsequent analysis and leave the less important plots in the appendix (to avoid distracting the readers).

3. Regression Analysis and Results (8 marks)

Build regression models to answer the research questions raised in the Introduction. Use model diagnosis and selection methods to reach a model that you are satisfied with. Report the results from the final model.

4. Discussions and Limitations (2 marks)

Discuss the possible issues with the data analysis. For example, are there confounding variables not included in the dataset? Do you get counter-intuitive results from the regression analysis? Do you have an explanation for these results? Any data quality issues, e.g., related to missing data, outliers, and high-leverage points? These are just a few examples. You can discuss more.

5. Interpretation & Conclusions for a Lay Audience (5 marks)

This section links back to the introduction and covers the interpretation of the results in more detail. It is the most important part of the report for non-statisticians (e.g., policy-makers). Because it must be easy to understand for a non-statistician, you need to write most clearly and in a non-technical way. Provide suggestions to policy-makers to improve the life expectancy for countries with different health and economic factors.

R code

Submit your R code as a separate file (as an R/RMD file) in your submission. The R code will not be marked. However, if you fail to submit your code, or if your code is not executable, then five marks will be deducted from your report.