

## Codebooks & Research Questions

The purpose of this assignment is to:

1. Read an example of research conducted using secondary data.
2. Begin thinking about possible research questions for your individual research project in statistics.

Assignment Preparation:

Download the following article from the Week 4 module on Canvas:

Anna D. Peterson & Laura Ziegler (2021). Building a Multiple Linear Regression Model With LEGO Brick Data, *Journal of Statistics and Data Science Education*, 29:3, 297-303, DOI: 10.1080/26939169.2021.1946450

Homework-

1. Read Anna D. Peterson & Laura Ziegler (2021), following these guidelines for reading scientific research papers:
  - **First, read the introduction** (skip the abstract – typically you would read the abstract to decide whether or not to continue reading the whole article). From the introduction you should be able to determine the population of interest, the authors' motivation for doing the study, the hypothesis, explanatory and response variables, and key findings.
  - **Second, read the Conclusion section.** This section will detail the key findings and implications of the study.
  - **Third, skim the Methods section.** Determine the sampling approach, sample size, and how the explanatory and response variables were measured.
  - **Fourth, skim the Results section.** Pay attention to tables and graphs, they are a good way to get an overview of the results.
  - **Lastly, return to the methods and results sections if more detailed information is needed.**
2. Answer each question below with one or two complete sentences:
  - a. In the context of the study, what is "complexity"? How was it measured?
  - b. What is the population of interest?
  - c. What associations tested in the study? What role type is/are that/they?
  - d. What is the main statistical method employed in the study?
  - e. What is the sample size used in the study?
  - f. Do you think correlation will be appropriate for the two variables in the scatterplot? Explain
  - g. Suppose, correlation is appropriate, describe the strength and direction of the relationship between the variables, using the scatterplot.
  - h. Comment on the variability in the dataset using Table 1.
  - i. With one or two paragraph summarize the key findings of the study.
3. Your instructor will share with you a number of datasets you can use for your research project in statistics. Choose one of the datasets that piques your interest, read information about the dataset online, and **write one to two paragraphs** summarizing what you have learned. Include information about how long the research has been ongoing, the objectives, population of interest, countries involved, funders, sampling and data collection methods, sample size, topic focus, etc.
4. Download (from Canvas) the codebook for the dataset you chose to review in question 3. Scan through the codebook from beginning to end. **List three topics** covered in the survey that you might find interesting for a research project in statistics. For example, if there are several questions in the codebook about poverty, then poverty is a topic covered in the survey, and therefore, could be a topic for your research. (Obviously, only topics included in one of the codebooks shared by your instructor can be selected for your research project in this course.)
5. One of the simplest research questions that can be asked is whether two variables are associated. For example:
  - Is medical treatment seeking associated with socio-economic status?

- Is water fluorination associated with number of cavities during dentist visits?
- Is alcohol use in adolescence associated with diagnosed mental health conditions in young adulthood?
- Is the age at which adolescents become sexually active associated with the choice of contraception?

Based on the topics you identified in question 4, **write three possible associations** you could test using the variables in the codebook. Use the template below.

IMPORTANT: Each association MUST be between two variables in the codebook! If one of the variables in the association is not in the codebook than the relationship cannot be tested!

- "Is \_\_\_\_\_ associated with \_\_\_\_\_?"
- "Is \_\_\_\_\_ associated with \_\_\_\_\_?"
- "Is \_\_\_\_\_ associated with \_\_\_\_\_?"

**ASSIGNMENT SUBMISSION:** Submit answers to questions 3, 4 and 5 only. Upload MS Word document to Canvas by 11:59 pm on Thursday, 16<sup>th</sup> February.