

# Data Analysis Assignment 1

COMM 7370 | Spring 2026

## Setup

- 1) Find the journal article below and download the data and codebook from OSF (link available in paper). In this assignment, you will recreate three of the variables (`mirth`, `need for humor`, and `engagement intentions`) from the article. We will use these variables to examine some differences and relationships.

Yeo, S. K., Su, L. Y.-F., Cacciatore, M. A., Zhang, J. S., & McKasy, M. (2023). The differential effects of humor on three scientific issues: Global warming, artificial intelligence, and microbiomes. *International Journal of Science Education, Part B*, 13(1), 59–83. <https://doi.org/10.1080/21548455.2022.2123259>

- 2) To start, load the following packages in R.

- `tidyverse`
- `summarytools`
- `rstatix`

- 3) Read the data into R. Answer the following questions as comments in your R script.

- a) How many observations are there in this dataset?
- b) How many variables are in the dataset?

- 4) How many respondents are at least 18 years old and consented to participate in the study (**Hint:** Look for the consent question in the codebook)? If necessary, create a new object in your R environment that contains only respondents who are at least 18 years old and consented to participate in the study.
- 5) These data come from an experiment—read about the experimental design in Yeo *et al.* (2023). Then, examine a frequency distribution of the variable, `issue`. Based on the frequency distribution of `issue` and the information in the Yeo *et al.* (2023), what does the variable `issue` describe?
- 6) Next, examine a frequency distribution of the variables, `gwhumor`, `aihumor`, and `mhumor`. Using the frequency distributions and the information from Yeo *et al.* (2023), explain what these variables, `gwhumor`, `aihumor`, and `mhumor`, describe.

- 7) Now you will create some new variables from existing columns in the dataframe. Create a new variable called `mirth` by averaging the existing variables, `Q86_1`, `Q86_2`, `Q86_3`, `Q86_4`, and `Q86_7`. Before creating the variable, read about the concept of mirth in Yeo *et al.* (2023). Each respondent in the dataframe should have their own score for `mirth`. Include commands in your R script to run a frequency distribution of `mirth`.
  - a) What is the mean value of `mirth` for the whole dataframe?
  - b) What is the mean value of `mirth` among respondents who were exposed to the issue of microbiomes?
  - c) What is the mean value of `mirth` among respondents who were exposed to the issue of artificial intelligence?
  - d) What is the mean value of `mirth` among respondents who were exposed to the issue of global warming?
- 8) Next, we will test whether the experimental conditions had an effect on mirth for the three issues.
  - a) Identify the dependent variable. Be sure to write this down as a comment in your R script.
  - b) Identify the independent variables (**Hint:** Refer to your answer to Question 6).
  - c) Identify and conduct the inferential statistical test(s) that you will use to determine whether there are differences in mean levels of `mirth` between the experimental conditions for each issue. **Note:** You should conduct three statistical tests, which will each result in a test statistic and a *p*-value. Be sure to note the test statistic and *p*-value for each test in your R script.
  - d) Are the mean levels of `mirth` statistically different for each of the three issues (**Hint:** Refer to your notes on inferential statistics to determine how to tell if your statistical test is significant)? Describe your results for each issue.
  - e) Re-create the individual graphs in Figure 3 in Yeo *et al.* (2023).
- 9) Now, we will create another new variable, `engage`. This variable captures participants' intentions to engage with the Twitter conversation (see Yeo *et al.*, 2023). `engage` is created by averaging participants' scores to the following three question items: (i) "I would like the original tweet," (ii) "I would retweet the original tweet," and (iii) "I would share the original tweet."
  - a) What are the columns that correspond to these three question items called in the dataframe (**Hint:** Look in the codebook for these statements)?
  - b) Create the variable `engage`.
  - c) What is the mean value of `engage`?
  - d) What is the standard deviation of `engage`?
- 10) Now, we want to know how mirth is related to participants' intentions to engage with the content they saw.
  - a) What statistical test might you use to examine this relationship? State this clearly in your R script.
  - b) Conduct the statistical test.
  - c) Describe the results of your test. Remember to report a test statistic and *p*-value. Are they significant? If so, what is the value of the test statistic? What does this coefficient mean?
- 11) Next, we want to create a variable that captures participants' *need for humor (NFH)*. This concept is a human trait that "defines one's tendency and inclination toward experiencing humor from others and being a source of humor or generating it" (Yeo *et al.*, 2023, p. 65).

- a) Based on the *Measures* section of Yeo *et al.* (2023), how many question items were used to construct the variable that measures need for humor (NFH)?
  - b) The 12 items used to create this variable are listed in the codebook as Q8 and Q9. Use these to create a new variable, NFH. Average the items together to create NFH.
  - c) What is the mean and standard deviation of NFH?
- 12) Form a hypothesis about how `mirth` and `need for humor` are related. Test your hypothesis using the appropriate statistical test. Be sure to report the results (test statistic, *p*-value) of your test. Describe your results and interpret what they mean.
- 

## Submission

Submit your R script (named `Assignment-#_FirstName-LastName.R`) to Canvas.

Your R script should:

- 1) Include commands and functions that are necessary to address all the questions in the assignment.
- 2) Contain comments that answer the questions in the assignment.
- 3) Run in its entirety without errors.

### ! Important

These standards apply to all submissions in this course that require R scripts. You should follow these instructions for preparation, naming, and saving of your R script for all of your individual lab assignments.