PSYC 11 Data sleuthing lab

Jeremy R. Manning

🕵 Overview

What can (and *can't*!) data tell us? In this lab, your team will explore this important issue by playing two roles: **data creator** and **data sleuth**. Your job in the "creator" role will be to (as the name implies) *create* a dataset to be shared with another team. Later, once you've passed off your team's dataset and received a new one from another team, your team will transition into a "sleuth" role, where you'll examine and analyze different aspects of the dataset.

🕵 Learning objectives

The goal of this lab is to learn about the "results" sections of scientific articles. In general, results sections report the *results* of the analyses described in the paper's methods section. A good results section will also help its audience to understand the basic logic of the study, any critical design decisions pertaining to the analyses, and how the different findings might be interpreted. To that end, you'll:

- Practice communicating clearly and directly
- Improve your understanding of what goes into creating a dataset
- Gain intuitions about what you can versus can't conclude from a given dataset
- Practice wrangling and analyzing data, creating figures, and running statistical tests
- Practice interpreting results
- Practice thinking about study design, resources, effort allocation, and time management

🕵 Procedure

We'll divide the class into three teams: A, B, and C. Each team will begin by finding or creating a dataset, following some specific guidelines (outlined below). You'll also create some "documentation" for the dataset. As part of putting your dataset's documentation together, you'll come up with a set of questions to be explored or answered about your dataset.

In the second phase of the lab, you'll pass your dataset onto another team and receive a dataset from a different team (A will give their dataset to B; B to C, and C to A). Using the "sending" team's questions as a guide, the "receiving" team will work with their designated teaching assistant to examine the dataset.

You'll conclude the lab by drafting a "results section" based on the dataset you analyzed.

🔎 Part 1: Construct a dataset!

Your team's dataset can be made-up (fake) or real (e.g., downloaded from a public repository, etc.). However, whether you create a fake or real dataset, it should be *interesting* (broadly construed), and it must follow several formatting guidelines:

1. Your dataset should contain a minimum of 500 observations and a maximum of 10,000 observations.
2. Each observation should include at least 5 features and no more than 100 features. For example, if you were constructing a dataset about fish swimming in an aquarium, each "feature" might be some aspect of a fish's position in the tank (e.g., x, y, and z coordinates, and perhaps the yaw, pitch, and roll of the fish relative to each axis).
3. You should save your dataset as a Google Sheet (set sharing permissions to "anyone with a link can view"):
    - Top row: column labels (single word or abbreviation, lowercase, no non-letter characters)
    - Each subsequent row should have a single observation
    - Each column should correspond to a single data feature
    - Data values should either be integers (1, 2, 3, etc.), real numbers (3.14, 2.71, -45.6789, etc.), dates or times (12:50:00 04/18/2022, etc.), or character strings ("happy", "sad", "angry", "excited", etc.; note: strings don't need to be enclosed by quotation marks).
    - The formatting should be consistent throughout the dataset (to facilitate analysis)
    - The formatting should be consistent within each column– e.g., all of the values in a column should have the same "type"
4. Create documentation for your dataset (save as a Google Doc; set sharing permissions to "anyone with a link can view"). Include the following information:
    - Where did the data come from? This can be made up, but it should be explained! Interesting back stories are encouraged. Help to motivate the team you're sending the dataset to.
    - How were the data collected? Briefly explain, in 1–2 paragraphs.
    - What do each of the features mean, in plain English?
    - What are the observations? (E.g., timepoints? People? Trials?)
    - A list of 5 questions:

  - All questions must seem plausible
  - All questions must be answerable (or potentially answerable) with a relatively straightforward analysis (e.g., PSYC 10-style statistical test, a basic chart or figure, etc.). In other words, you should keep the scope of the questions relatively narrow and specific.
  - At least 1 question should be possible to explore/answer using the dataset
  - At least 1 question should be *im*possible to explore/answer using the dataset
  - All other questions can either be possible or impossible to explore/answer using the dataset
  - Keep track of which questions can/can't be answered with the dataset, but keep those labels hidden (i.e., don't put it in your to-be-shared documentation)

Once you've created your team's dataset, documentation, and "possible vs. impossible" labels for each question, upload your materials using this form.

🔎 Part 2: Check out the data!

Each team will analyze data from their designated source team. Answer each of the five questions included in your received dataset's documentation using any approach you deem appropriate. Figures and stats are encouraged!

For each question, document the following:

- State any assumptions you're making to answer the question
- Explain how you'll approach answering question with the given analysis. In other words, explain how what you're trying to discover relates to the analytic tools and approaches you're employing.
- Explain what conclusions can or can't be drawn from the given analysis (according to how the analysis turned out).
- If you think a particular question is impossible to answer, explain why– and what you'd need (from the dataset, analytic tools, etc.) in order to be able to answer that question. You should still carry out *some* sort of analysis for each question; the idea is to notice and document when the insights possible to attain from a particular analysis fall short of the question(s) you're trying to answer about the data.

Your team can share a common document and set of analyses, figures, stats, etc.

🕵 Writing your lab report

Your writeup for this lab will mimic a "results section" of a scientific article, about the dataset you analyzed with your team. You should include the following elements:

- For each question:
  - Create at least one figure
  - The figure should include a caption describing what the figure is showing

  - Also write (roughly) a paragraph of text describing the key analyses, results, relevant interpretations, etc. For example, describe "answers" or "insights" into the question that you gained through your analyses, or describe how the question is impossible to answer given the data, etc.
- Organize the questions and paragraphs so that they tell a "story" about the dataset.
- Add transition sentences as needed.
- Add an "overview" paragraph at the beginning to summarize the dataset, how it was analyzed, and what you found.

🕵 Closing discussion points

The results section of a scientific paper is, in many ways, its core. The "merit" of a paper typically rests on how accurately and effectively it communicates the main findings of the study. For example:

- Was the dataset appropriate for answering the proposed questions?
- Were the analyses appropriate?
- Were the interpretations and/or conclusions justified by the data (or analyses)?
- Are the results visualized in an intuitive way?
- Are the results communicated clearly?
- Was the "logic" of the analytic approach clear and/or easy to follow?

Achieving these ideals is not simply a matter of listing sets of numbers or the outcomes of a series statistical tests. You need to help your intended audience understand *why* the approaches you took helped you to understand the questions you explored, how the questions fit together into a cohesive narrative, and so on. In other words, you need to make your "story" easy to read, easy to believe, and easy to follow.