Understanding the Research Process

Dr Austin R Brown

Kennesaw State University

Introduction

▶ When we hear the word "science," we probably think of people in lab coats surrounded by beakers and test tubes performing experiments of some kind.

  ▶ Some of you may have done this type of work!

▶ But we can define science much more broadly.

  ▶ Science can be generally defined as a *systematic* and *organized* way of acquiring knowledge about the natural world, its phenomena, and the principles governing those phenomena.

Introduction to the Scientific Method

▶ Part of the systematic and organized way we go about performing science is laid out for us by the scientific method (yes that same scientific method from high school science class).

▶ The steps in the scientific method are:
1. Observation
2. Question
3. Hypothesis
4. Experiment
5. Data Analysis
6. Conclusion
7. Communication

▶ Let's consider a scenario and walk through each of these steps together!

Observation: Research Scenario

▶ I am old enough to remember a time before social media. When it came into prominence when I was late in high school and early in college, I couldn't have imagined what it would grow into in such a short amount of time. It is amazing to me how easy it is to connect and share knowledge with people all over the world. Of course, there are problems with social media becoming so intertwined with our lives. I have especially noticed (or observed) people I know who spend lots of time on apps like Instagram exhibiting depressive symptoms. So that led me to wonder if the amount of time people spend on social media is associated with or related to depressive symptoms.

Question: Formulating the Research Question

▶ So I have observed something in the world that I'm interested in studying further. This is a great first step! Now, the next step is to formulate the research question.

▶ The research question is key to the scientific method because it lays out the "roadmap" of our study including:

  ▶ The specific group or groups we are interested in studying, which I call the *Population(s) of Interest* (the "who")

  ▶ The specific qualities (i.e., variables) about the group or groups we are interested in studying further, which I call the *Variable(s) of Interest* (the "what")

  ▶ The quantitative summary characteristic of the variables we are specifically interested in studying further (e.g., mean, median, proportion, slope, etc.), which I refer to as the *Characteristic of the Variable to be Analyzed* (more of the "what")

  ▶ The *General Analytical Method* we will be using to perform the analysis (e.g., comparison, association) (touches on the "why")

Question: Formulating the Research Question

▶ It is of critical importance to do this at the onset of the study, before data has been collected to ensure that we are getting the right data to answer our question!

▶ So in our case, what would the elements of our research question specifically include? We have decisions to make!

Question: Formulating the Research Question

▶ **Population of Interest:** Suppose I'm specifically interested in University of Georgia undergraduate students. This is my population or group that I'm going to be collecting data from.

▶ **Variables of Interest:** Here I have two variables which I need to more clearly define: time spent on social media and depressive symptoms.

▶ Let's pivot for a moment to talk more about the important topic of variables.

Variables

▶ First of all, we have all heard the term "variable" in a math setting, but what does it mean in a scientific setting?

▶ In science, a **<u>variable</u>** is some characteristic of something.
  ▶ Hair color is a variable
  ▶ Number of minutes per week spent on TikTok is a variable
  ▶ Monthly amount of money spent on Dunkin coffee is a variable

Variables

▶ Okay, so we can see that variables can be further classified into more specific types of variables.

▶ A variable which is inherently measured by numbers is called a **quantitative** variable.

　▶ Number of minutes per week on TikTok and monthly amount of money spent of Dunkin coffee are both examples of quantitative variables.

▶ A variable which cannot be inherently be measured by numbers (like hair color) is called a **categorical** or **qualitative** variable.

　▶ Customer response surveys which ask us questions like "How satisfied were you with your Dunkin coffee on a scale from 1 - 5 with 1 meaning totally dissatisfied and 5 meaning totally satisfied?" are also categorical variables because the numbers are representing the customer's qualitative opinion!

Variables

▶ We can even further differentiate between quantitative variables.

▶ If a fraction or decimal value makes sense for a quantitative variable, then we call this a **<u>continuous</u>** quantitative variable
  ▶ Number of ounces in a can of Coke
  ▶ Amount of time it takes to commute to campus
  ▶ Temperature on the surface of the Sun

▶ If a fraction does not make sense for a quantitative variable, then we call this a **<u>discrete</u>** quantitative variable
  ▶ Number of Coke cans in my fridge
  ▶ Number of undergraduate students enrolled at KSU
  ▶ Number of free throws Trae Young makes this year

▶ Okay so circling back to the research question, we have two variables which we need to better define.

▶ Maybe I can measure social media time by asking the UGA undergraduate students to show me their daily average usage of their top social media applicaiton (mine is suprisingly Instagram!). So our variable is not just social media usage, but it is specifically daily average amount of time spent on the most used social media application.

    ▶ This is an example of a quantitative, continuous variable.

Question: Formulating the Research Question

▶ What about depression? Typically, depression is measured using rigorously tested surveys developed by psychologists. Once a person fills out a survey, their responses are summed into one total numeric score with greater values generally denoting more severe depressive symptoms and vice versa.

　　▶ So suppose we choose a survey we want to give to the UGA undergraduate students. At the conclusion of the survey, we will end up with a quantitative score. So the variable "Depression Severity" is also a quantitative variable, but most likely a discrete quantitative variable.

Question: Formulating the Research Question

▶ Next, we want to specify the characteristic to be analyzed. Here, what I'm truly most interested in is determining if there exists an association between social media usage and depression. Thus:

▶ **Characteristic to be Analyzed:** Correlation (we'll talk more about this later in the semester)

▶ **General Analytical Method:** Association

Question: Formulating the Research Question

▶ Putting it all together:

"Is there evidence that an association may exist between average amount of time spent on social media per day and depressive symptoms among University of Georgia undergraduate students?"

Hypothesis: Specifying the Research Hypothesis

▶ Wonderful start! Now that we have formulated our research question, the next step in the scientific method is to formulate a *hypothesis*, or educated guess, as to what the answer to our research question will be.

▶ Based on my experience and review of relevant literature, I hypothesize that in our study, we will find evidence for an association between average amount of time spent on social media per day and depressive symptoms among University of Georgia undergraduate students.

▶ This serves as my *Research Hypothesis*

Hypothesis: Specifying the Statistical Hypotheses

▶ After we specify the research hypothesis, the next step is to specify the statistical hypotheses.

▶ We will discuss what these hypotheses are and how to specify them in a later lecture after we have a brief conversation on probability.

Experiment: Collecting Data

▶ Now that we have set up our hypotheses, we can then begin collecting data to help us determine whether evidence exists which supports or does not support our research hypothesis.

▶ When we collect or sample data, we are gathering this information from our population or populations of interest.
  ▶ So in our case, our population is UGA undergraduate students.

▶ Ideally, we would be able to ask every single member of our population to be in our study!
  ▶ But realistically, this isn't feasible.

Experiment: Collecting Data

▶ In real research, we almost never are able to work with every member of the populations we're interested in because of practical constraints.

▶ So instead, we work with smaller subsets of the population called *samples*.

▶ Recall that the goal of the study is to answer our research question: do we have evidence for an association between social media usage and depression score?

  ▶ I'm interested in answering this for the whole population.

▶ However, I'm limited in my ability to fully answer this question because I am unable to gather information from every member of the population.

  ▶ How do we rectify this issue?

Experiment: Collecting Data

▶ The answer? Through *random sampling*.

▶ When we perform one of the random sampling techniques, we are attempting to include all of the diversity that exists in our population in our sample data.
  ▶ So, for example, making sure the proportion of freshman, sophomores, juniors, and seniors in our sample is roughly equivalent to those same proportions which exist at the population level.

▶ We often are not be able to do a perfect job of this, and when this is the case, we need to make note of it as this changes to whom conclusions can be drawn.

▶ How do we perform random sampling? Let's talk through a few common techniques!

Experiment: Collecting Data

▶ **Simple Random Sampling** - the most common type of sampling method. Here, each member of the population is equally likely of being included in our sample.

  ▶ So here, we could put every student's name on a little piece of paper, put them in a hat, and pull out some desired number of pieces, and go ask those specific students.

▶ **Systematic Sampling** - numbering a population and including every kth member in the sample.

  ▶ Here, we could stand outside of the Student Center and ask every 10th student to respond to our survey questions.

Experiment: Collecting Data

▶ **Stratified Sampling** - taking a larger population and breaking it into smaller, heterogeneous groups.
   ▶ For example, we could break down the UGA undergraduate population by class (e.g., freshman, sophomores, etc) and then take simple random samples from those groups.

▶ **Cluster Sampling** - the population is divided into clusters, and a random selection of clusters is chosen for inclusion in the study, with all individuals or elements within the selected clusters becoming part of the sample.
   ▶ So here, maybe our clusters are classes (like STAT 3125 Section 1). We randomly select some number of classes and then all of the students who attended class that day are included in the sample.

Experiment: Collecting Data

▶ You may be asking yourself: "So when do we perform the experiment?"

▶ We already did through data collection. The act of asking random students how much time, on average, they spend on their top social media app per day and asking them to fill out the depression survey yields random albeit well-defined outcomes.

▶ By definition, we have performed an experiment!

Data Analysis

▶ Now that we have collected our data, we can move into data analysis!

▶ In other words, we are taking raw data and synthesizing it using methods specific to our variable types and research question to determine whether we have substantial evidence supporting or not supporting the research hypothesis.

▶ I refer to these methods as *evidence gathering techniques*.

Data Analysis

▶ Generally speaking, we have two types of evidence gathering techniques:

▶ **Descriptive Techniques:**
  ▶ Data Visualizations (e.g., a bar plot or scatterplot)
  ▶ Quantitative Summaries (e.g., frequency, proportion, mean)

▶ **Inferential Techniques:**
  ▶ Hypothesis Tests
  ▶ Confidence Intervals
  ▶ Effect Sizes

Conclusion

▶ Once we have performed our data analysis, we have two questions to answer:

1. Do the data provide substantial *statistical* support for the research hypothesis?
2. Do the data provide substantial *practical* support for the research hypothesis?

▶ We will learn later in the semester that sometimes we have different answers to these two questions.

Communication

▶ Now that we have completed our study, it is now time to effectively communicate the results in a reproducible manner.

▶ Often, this is done through a written medium via academic manuscripts published in academic journals.

▶ To write one of these manuscripts, we need to make sure we are as detailed as possible (without violating confidentiality etc.) about exactly what we did to conduct our research.

Communication

▶ In general, all manuscripts contain the same elements: (1) Introduction, (2) Data Description, (3) Analysis & Results, and (4) Conclusions & Future Directions.

▶ So what does each of these sections entail?

Communication: Introduction

▶ In the introduction, you lay out the "who", "what", and "why". More specifically:

▶ **Who** is being studied (or what is the population(s) you're interested in)? Be specific about timing of data collection as well if that's relevant!

▶ **What** about the population is being studied or what are the variables of interest)?

▶ **Why** does this matter or to whom would the results be most interesting?

▶ Basically, you're writing out the research question in a paragraph format.

Communcation: Data Description

▶ Here, you tell the reader more about the variables themselves.

▶ Restate the variable names. What is your outcome variable? Which variable(s) is/are your explanatory variables? Are they categorical or are they quantitative? What are the units of measurement (if relevant).

Communication: Analysis & Results

▶ In the analysis and results section, you tell the reader what descriptive and inferential methods you chose to use to answer the research question.

▶ Then, you include the results of those methods in the body of the document along with contextual conclusions/interpretations about what they mean as it relates to the research question/hypothesis.

▶ It is best to be thorough in this section with you descriptions of the methods used because a primary goal of writing this manuscript is reproducability.
  ▶ In other words, I should be able to perform an incredibly similar study to the one you did using your manuscript as a roadmap.

Communciation: Conclusions & Future Directions

▶ In this section, you write your conclusions. So you'll include: restating the purpose of the study, restating the conclusions you drew from the analysis, talk a little bit about the importance of the findings (remember, to whom are the results most important?), and then you'll talk about future directions.

▶ Often in research, the results of a study produce more questions than answers. So for example, suppose the results of our UGA social media study showed substantial evidence that there may be an association between social media usage and depression.

▶ Well one thing I would want to know is: what apps are being used the most by those with the most severe depressive symptoms and what are the characteristics of app usage?
  ▶ This would be a future direction.

Final Thoughts

▶ Performing science and conducting research is more than just data analysis.

▶ It is a series of systematic, methodical steps we take to ensure that we are properly answering our research questions.

▶ As we work through this semester, we will see how we work through these steps in different research contexts and you will have an opportunity to complete a manuscript like the one described previously.