Introduction to Experimental Design

Dr Austin R Brown

Kennesaw State University

Introduction

▶ When we think about statistics and data science, very often our minds go straight to the analysis of data.

▶ However, being able to effectively answer questions with data requires a systematic, well-thought-out approach.

▶ For us, the scientific method lays out a nice framework that we can use to guide our thinking.

Introduction

▶ Remember from high school science class that the scientific method generally involves:

1. Making a hypothesis about some phenomenom
    ▶ This includes defining our independent and dependent variables
2. Collecting data to test the hypothesis
3. Analyzing the data
4. Drawing conclusions from the data
5. Refining the hypothesis and repeating the process

▶ Even if we aren't working in laboratory sciences, this systematic approach helps us make sure we're using the right data to answer the right question.

Introduction

▶ I like to call working through steps 1 - 5 of the scientific method a **study**.

▶ Generally speaking, we can classify studies into two categories:
   1. **Observational Studies**
   2. **Experimental Studies**

Introduction

▶ In an **observational study**, we are simply observing our independent and dependent variables. We don't have control over how observational units get assigned to specific values of either the independent or dependent variables.

Introduction

▶ For example, suppose I wanted to know if the mean annual income differs between data science majors and psychology majors.

▶ In this case, mean annual income serves as my quantitative dependent (or outcome) variable and major (data science or psychology) serves as my categorical independent (or predictor/explanatory) variable.

Introduction

▶ I as the researcher in this case don't have control over whether students are data science or psychology majors – I'm simply observing a phenomenon.

   ▶ So this study would be classified as an observational study.

▶ Let's contrast this with an experimental study.

Introduction

▶ Let's say you work for a marketing department in a retail company. We have an email list of 10,000 customers.

▶ We want to test the impact of two different email subject lines (e.g., a generic subject line and a personalized subject line) on annual spending with our company.

Introduction

▶ In this case, we could randomly assign our customers to either the generic subject line group or the personalized subject line group and follow their spending over the course of a year before making a comparison.

▶ Email group serves as our categorical independent variable and annual spending serves as our quantitative outcome variable.

▶ But notice in this case, we the researcher assigned the participants to their respective groups.
  ▶ This is the key difference between observational and experimental research.

Introduction: Why Should I Care About Experimentation?

▶ While we may often associate experiments with laboratory science, experimental design is actually very helpful in a variety of fields including:
1. Engineering
2. Manufacturing/Quality Control
3. Marketing
4. Social Science
5. Educational Research
6. Much more!

▶ The roots of DOE (design of experiments) go back to Fisher himself as he aimed to study crop yields.

Introduction: Why Should I Care About Experimentation?

▶ Over the course of the semester, you will see how DOE is important to:

▶ **Informed Decision-Making**: It provides a structured approach to testing hypotheses, allowing people/organizations to make data-driven decisions based on reliable evidence.

▶ **Resource Optimization**: By identifying what works and what doesn't, people/organizations can allocate their resources more efficiently, avoiding wasted time and money on ineffective strategies.

Introduction: Why Should I Care About Experimentation?

▶ **Process Improvement**: Well-designed experiments can uncover insights that lead to innovative solutions and improvements, which may serve as a competitive advantage in a business setting.

Definitions

▶ Before we go much further, it may be helpful if we have some agreed upon definitions (so we're speaking the same language!)

▶ Note, if there is ever a time when a term is used in these slides or elsewhere that doesn't seem well-defined, **PLEASE ASK FOR GUIDANCE!!**

Definitions

- **Experiment (or Run)**: an action where the experimenter changes at least one of the variables being studied and then observes the effect of the action.
  - Randomizing our customers into the generic and personalized email groups and then observing their purchasing behavior was an experiment.

- **Experimental Unit**: the item under study upon which something is changed. This could be raw materials, human subjects, or just a point in time.
  - Our individual customers in the marketing example were our experimental units.

Definitions

▶ **Independent Variable (Factor or Treatment Factor)**: We generally think of this as the X variable that is being controlled at some level during any given experiment.

  ▶ Email group from the prior example

▶ **Lurking Variable (Background Variable)**: a variable that the experimenter is unaware of or cannot control which could have an effect on the outcome of the experiment.

  ▶ In the email example, annual income probably plays a role in annual spending. This isn't something the marketing department can control.

Definitions

▶ **Dependent Variable (or Response or Outcome)**: Usually denoted Y, this is the characteristic of the experimental unit that is measured after each experiment/run.

▶ **Effect**: The change in the response that is caused by a change in a factor/independent variable.

▶ These definitions will get us started, but there will be more new terms added as we progress through the semester!

Planning Experiments

▶ The key to successful experiments (and studies in general really) is a very clear articulation about what you're studying, why you're studying it, how you're studying it, and how you'll draw conclusions from the experiment(s).

▶ More specifically: **(1)** Define the objective **(2)** Decide what the outcome is (and how it will be measured) **(3)** Determine the independent variables and possible lurking variables **(4)** Choose the design (more on this as the semester progresses) **(5)** Be clear on data collection processes/procedures **(6)** Be clear on which analyses will be performed **(7)** Draw conclusions