STAT 514
Design Of Experiments

Lin Wang
Department of Statistics
Purdue University

Grading Policy and Exam Schedule

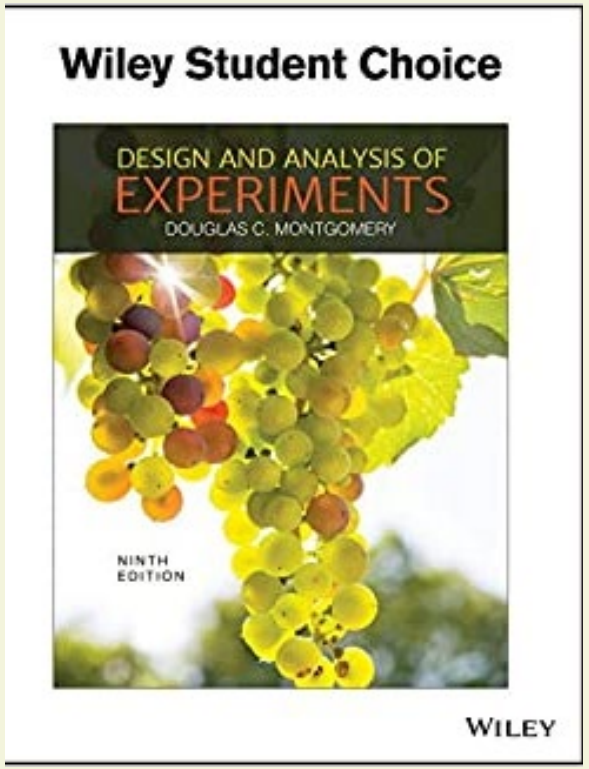
- Midterm exam 1 (15%)
- Midterm exam 2 (15%)
- Final exam (30%)
- Final project (20%)
- 8-10 Homework assignments (20%. The lowest score will be dropped.)
- Exams are open book.

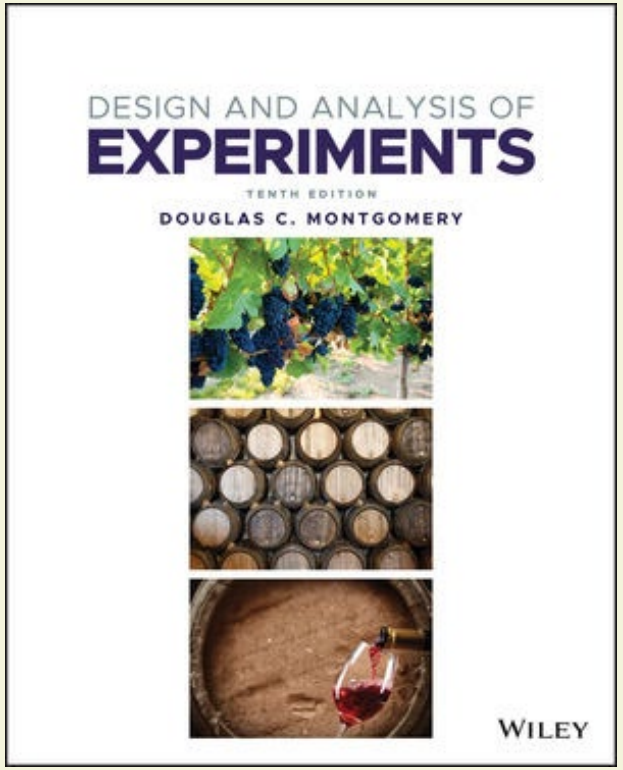
Prerequisites

- Linear algebra is required. Familiarity with the notation of matrices is assumed.
- An introductory course in statistics is required.
- R coding is required for computation and homework assignments.

You do not have to be proficient in R at this time, but be prepared to learn and use some fundamental coding techniques in this course.

The Book



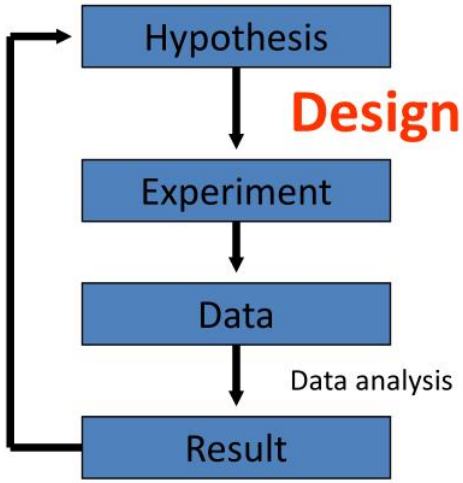


- Either edition is good to use
- Homework assignments may from the book and other resources

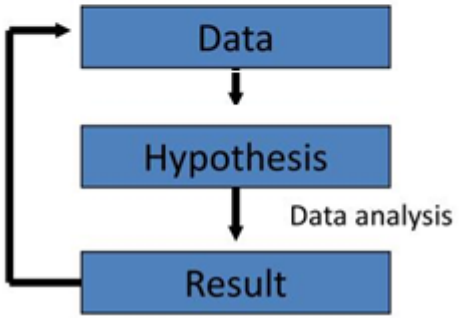
Why Conduct an Experiment?

Answer: To collect data.

Hypothesis Driven Research



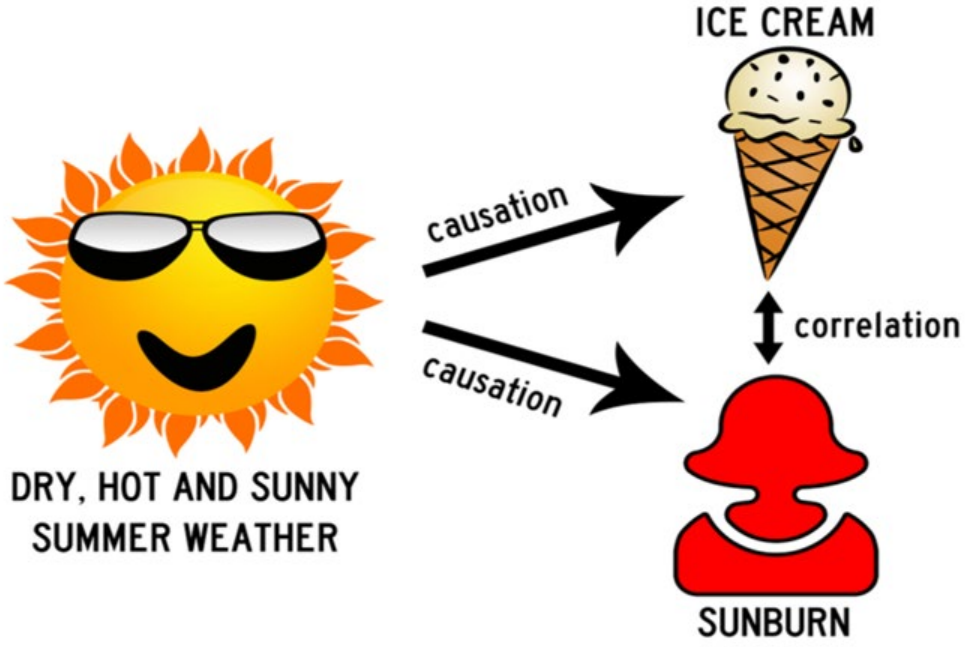
Data Driven Research



Experimental data: small data can provide a reliable result
Mini example: In an e-commerce site, customers who see an ad buy more.
Observationally this could be due to targeting. A randomized trial shows the true factors.

Objectives of Experiments

1. Removing confounding (Correlation vs Causation)



- The dry and hot summer weather is a confounder.
- Confounder brings issues
- A confounder (or 'confounding factor') is something, other than the thing being studied, that could be causing the results seen in a study.
- Confounders have the potential to change the results of research because they can influence the outcomes that the researchers are measuring.
- Experiments can avoid these issues and help us make reliable conclusion on causal effect

Another Case study: Vitamin supplement users appear healthier in claims data. After randomization there is little effect because supplement use is confounded by health-seeking behavior.

Takeaway: Experiments remove confounding by randomization and control.

Objectives of Experiments

2. Screening factors

- **Goal:** Identify which inputs matter.
- **Example A: PCR yield in a biotech lab**
- Response: DNA yield
- Candidate factors: magnesium ion (Mg^{2+}) concentration (low vs high), annealing temperature (low vs high), primer concentration (low vs high), cycle number (low vs high)
- **Design:** factorial design
- **Outcome:** Main effects screen; possible interactions

- **Example B: Semiconductor line yield**
- Response: percent of wafers passing
- Factors: photoresist thickness, bake temperature, exposure time, developer concentration, spin speed
- **Design:** factorial design

Objectives of Experiments

3. Optimizing the process

- **Goal:** Find the settings that maximize or minimize a response.
- **Example: Bioreactor titer optimization**
- Response: titer (g/L)
- Important factors from screening: temperature and pH
- **Design:** Central Composite Design for active/ sequential learning
- **Alternative example:** Coffee taste score vs grind size and brew temperature with a CCD.

Objectives of Experiments

4. Saving experimental cost

- Example: The design of a business' web page has potentially important economic impact. Suppose that the website has the following components:

(1) a photoflash image (4 choices), (2) a main headline (8 choices),

(3) a subheadline (6 choices), (4) a main text copy (5 choices),

(5) a main image on the right side (4 choices),

(6) a background design (3 choices), (7) a footer (7 choices).

- We are interested in finding the factors that influence the click-through rate.

A total of $4 \times 8 \times 6 \times 5 \times 4 \times 3 \times 7 = 80{,}640$ web pages, not feasible!

We will study experimental design strategies to save the number of web pages that have to be considered in the experiment.

Objectives of Experiments

5. Product Evaluation

- **Example: Vaccine efficacy trial**
- Randomly assign 30,000 participants per arm to vaccine or placebo
- Outcome: symptomatic infection within 3 months
- **Analysis concept:** risk ratio and its confidence interval

Objectives of Experiments

6. **Handling nuisance variables**

- **Goal:** Increase precision by controlling the variation of nuisance variables

- **Example A: Cookie baking competition**
- Block by oven to remove oven-to-oven variation
- Randomize run order within each block
- **Example B: Multi-center clinical trial**
- Block by site or stratify by age group to remove known sources of variation.