Mathematics 231

Lecture 13

Liam O'Brien

Announcements

- Reading
  - Today M&M 3.1 178-191
  - Next class M&M 3.2 197-207

Topics

- Collecting data
  - Sampling
  - Observational studies
  - Experiments

What is Statistics?

- Collecting Data
- Describing Data
- Drawing Conclusions from Data

Collecting Data

- Production of data requires answers to the following questions:
  - Who or what is the object of the study – units?
  - What variables should be measured?
  - How should we select the individuals/groups/countries?

Collecting Data

- Statistical designs for producing data usually rely on either **sampling** or **experiments**.
- **Sampling**: Basic idea of sampling is to study a "part" to gain information about the "whole."
- Examples:
  - U.S. opinion polls often use only 1000-1500 people.
  - U.S. unemployment surveys use approximately 50,000 households.

Sampling

- Reasons for sampling:
  - Cost: Too expensive to “sample” everyone.
  - Timeliness: Too time-consuming.

- However, to ensure sound conclusions the sampling design is very, very important.

Observational Study vs. Experiment

- A sample survey collects information on a population, leaving it undisturbed by the process.
- Sample surveys are a type of **observational study**.
- In an observational study there is no attempt to influence the responses.

Experiments

- An experiment imposes some "treatment" on the individuals in order to observe their responses.
- An experiment allows us to control or eliminate lurking variables.
- In principle, experiments are the "gold standard" of evidence to support "causation."
- Experiments may not always be ethical/practical.

Example: Study of Welfare

- In a large observational study, children on welfare were found to have behavioral problems later in life.
- However, the effect of being on welfare on children's behavior is confounded with the socioeconomic status of the families that need welfare.
- In principle, an experiment can control or eliminate these effects.
- Is this ethical/practical?

Observational Study vs. Experiment

- In an observational study there is no attempt to influence the responses.
- An experiment imposes some "treatment" on individuals in order to observe their responses.
- An experiment allows us to control or eliminate lurking variables.

Example: Pheromones

- **Pheromones**: Chemicals that regulate behavior in many animals.
- Pheromones secreted by one animal affect behavior of another animal (e.g., mating, finding food, escaping enemies).
- Is there any evidence of human pheromones?

Example: Human Pheromones

- Martha McClintock – Wellesley Class of 1969

Example: Human Pheromones

- McClintock 1971 study.
- McClintock conducted a study of 135 women in her dorm; wrote up the results as senior thesis.

- Later published in *Nature*, considered the first scientific evidence of functioning of human pheromones.

McClintock 1971 Study

- Recorded menstrual period onset dates of students in Wellesley dorm in October.
- By March, found close friends and roommates had onset dates that were closer in time.
- Evidence for pheromones in humans?
- Confounding variables?

McClintock 1998 Experiment

- Researchers gathered samples from 9 women by placing pads under their armpits for 8 hours, taken during 2 different phases of the menstrual cycle.
- Pads frozen in glass vials, then swiped under the noses of 20 women.
- Depending on when in the cycle the samples were taken, they could shorten or lengthen cycles in other women.

Design of Experiments

- Terminology
- The **units** are the subjects of the study.
- A specific experimental condition is called a **treatment**.
- The distinction between response and explanatory variable is usually clear.
- Explanatory variables are often called **factors** and specific **treatments** are called **levels**.
- Note: Specific **treatments** can be formed by combinations of different factors.

Example: Physicians Health Study

- **Units**: 21,996 male physicians
- **Factors**: 2 factors

  Aspirin vs. placebo

  Beta carotene vs. placebo

  → 4 (treatment) levels
- **Response**: Heart disease/cancers

Factor 1: Aspirin	Factor 2: Beta carotene — Yes	Factor 2: Beta carotene — No
Yes	Aspirin, Beta carotene	Aspirin, Placebo
No	Placebo, Beta carotene	Placebo, Placebo

Design of Experiments

- **Single Treatment Experiment**
- What if we assign all units to receive the treatment? What can we say about the response?
- We need a **control group** to compare to.
- Control is the 1st principle of experimental design.

Control Group

- **Control group** differs from the treatment group only in terms of the treatment assignment.
- It allows us to eliminate or reduce the effects of experimental arrangements, subject selection, and placebo effects.
- If these are not controlled, the result is **bias** (i.e., study systematically favors certain results.

Placebo Effects

- Placebo effect is the measurable, observable, or felt improvement not attributable to treatment.
- Example: Patients suffering pain after wisdom tooth extraction got just as much relief from a fake application of ultrasound as from a real one (provided both the doctor and patient thought the machine was on).

Assignment to Treatments

- The 2nd principle of experimental design concerns assignment to treatments.
- How to make assignments in a way that's fair to all treatments.
- **Matching**: We can match someone who receives one treatment to someone who receives another based on important factors (e.g., age, gender, smoking status).
- **Randomization**: Use chance to decide.

Randomization

- Randomization produces treatment groups that are similar in all respects except treatment received.
- So differences in the response must be due to the treatment differences or to chance.
- The effects of chance should average out with a large enough sample size.
- A treatment effect so large that it would rarely occur by chance is said to be "statistically significant."

Principles of Experimental Design

1. Control effects of lurking variables by comparing two or more treatments.
2. Use randomization to assign units to treatments.
3. Replicate each treatment on many units to reduce chance variation.

Blinding

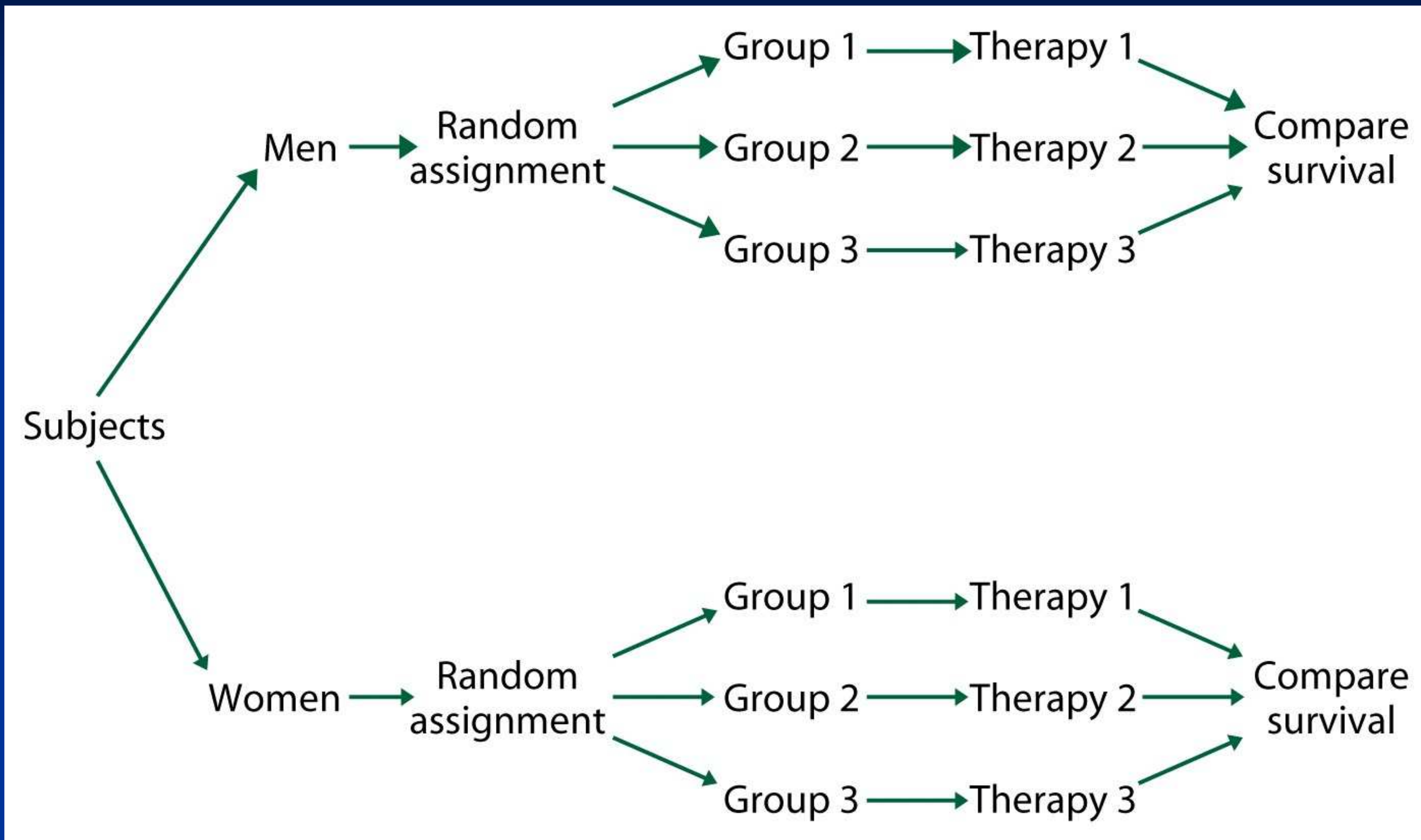
- **Blinding**: Comparison of treatments can be distorted if subjects, persons administering or evaluating treatment know which treatment if being allocated.
- Blinding avoids many sources of unconscious bias.
- **Double-blind**: Neither subjects nor experimenters know which treatments have been received.
- **Single Blind**: Subjects do not know which treatment they have received.

Blocking

- **Blocking**: A block is a group of units known to be similar in some way that is thought likely to influence the response variable.
- In a "randomized block design," randomization is carried out separately within each block.
- Example; **Matched-pairs design**
  - Blocks consisting of two units matched as closely as possible, e.g., one active eye drop in one eye and a placebo in the other.

Block Design

- Consider an experiment designed to compare three chemotherapies for a type of cancer that progresses differently in males and females.
- **Units**: 60 subjects (30 males, 30 females)
- **Factor**: Therapy (3 levels)
- **Block**: Gender (blocks consist of males and females)
- **Response**: Survival


Subjects
Men
Random assignment
Group 1
Group 2
Group 3
Therapy 1
Therapy 2
Therapy 3
Compare survival
Women
Random assignment
Group 1
Group 2
Group 3
Therapy 1
Therapy 2
Therapy 3
Compare survival

Choice of Blocks

- Blocks should be chosen on the basis of the most important (known) unavoidable source of variation among experimental units.
- Randomization then averages out the remaining sources of variability to allow unbiased estimation of treatment effects.
- Blocks allow greater precision, because a source of systematic variation is removed (reduced variability) from the experimental comparison.