STAT 3119 ANOVA

08/27/2019 @GWU

SYLLABUS: STAT 3119 (ANOVA)

- **INSTRUCTOR**: Xin Tian, Ph.D.
- **Semester**: Fall 2019 (15 weeks: Tues 08/27- Thurs 12/5)
- **Time**:
  - Tues/Thurs 6:10-7:25 pm (lecture); 7:30-8:00pm (Q&A)
- **Location**: Gelman Library Room #608
- **Office hours**: TR 7:30-8:00pm (Gelman Library #608)

Course Overview

- Textbook:
  - https://create.mheducation.com/shop/#/catalog/details/?isbn=9781307495218

    Applied Linear Statistical Models (5th ed.), by Micheal Kutner et al.: (customized e-book or print).
  - https://github.com/npmldabook/Stat3119

    My Lecture notes (posted online during/after the class)
- Learning outcome
  - **1. Learn the main methods of experiment design**
  - **2. Design your own experiment to gather data to answer your research questions**
  - **3. Apply the appropriate models and techniques to analyze and interpret the experiment data**

Class Materials Repository

- Organized by weeks
- Each weekly folder contains class lecture notes, code and data sets for examples in the corresponding chapter and the homework problems

Expected work and Grading

- Each week of the 15-week semester
  - **In-class: 100 mins lecture + 50 mins discussion (Q & A)**
  - **Out of Class: a minimum of 5 hours of independent learning and homework time (GW's credit hour policy)**
- Grading
  - **Homework/Quiz/Class attendance (20%),**
  - **Midterm exam (30%),**
  - **Final exam (40%),**
  - **Project (10%): a personal experiment or a summary of scientific paper**

Class Policy

- No smart phone use is allowed in the classroom.
- Students are expected to come to each class, and any planned coming late or leaving early to/from the class must be discussed prior to the planned schedule.
- Students should notify the teacher during the first week of the semester of their intention to be absent from class on their day(s) of religious observance.
- There will be NO make-up quizzes/exams.

Outline Week 1 (Chapter 15)

- A1. Introduction of Design of Experiments
- A2. Overview of Basic Concepts of DOE
- A3. Software Setup and Data Summary
- B1. Design of Observational Studies
- B2. Overview of Basic Statistical Concepts
- B3. Software Demo of Basic Analysis

Introduction of DOE

- Statistics is often taught as though the design of the data collection and the data cleaning have already been done in advance.
- However, as most practicing statisticians quickly learn, typically problems that arise at the analysis stage, could have been avoided if the experimenter had consulted a statistician before the experiment was conducted to collect the data.
- A well designed studied is usually simple to analyze and interpret. This course is to provide an understanding of how experiments should be designed so that when the data are collected, these shortcomings are avoided.

Overview of DOE

- Correlation is not causality. You've probably heard that before in any regression class. If you want to infer causality from data, then the best way is to use randomized experiments.
- In experimental design we look at how to choose the data that we will gather. In addition to being able to make causal conclusions, we also look at how to maximize the statistical efficiency of the generated data set.

Motivating Example

-

We have all done certain experiment at school, e.g Chemstry lab, or from our own experience such as baking/cooking.

- **How many of you have baked a cake? What are the factors involved to ensure a successful cake? Factors might include baking time, ingredients, baking temperature, etc ? You probably follow a recipe, i.e. someone did the experiment in advance!**

- **What parts of the recipe did they vary to make the recipe a success? Probably many factors, temperature and moisture, various ratios of ingredients, and presence or absence of many additives. Now, should you change several factors altogether? should you keep all the factors involved in the experiment at a constant level and just vary one to see what would happen?**

Brief History of DOE

- An early record of a designed experiment in the medical field is the study of scurvy by James Lind on board the Royal Navy ship Salisbury in 1747. Lind conducted a study to determine the effect of diet on scurvy and discovered the importance of fruit as a preventative measure.
- Modern experimental design as a subject is about 100 years old. The methods in this course date back to agricultural field trials. Since then the ideas have seen use in medicine, manufacturing, quality control, computer aided design and electronic commerce.

Eras of DOE

- The agricultural origins, 1918 – 1940s
  - R. A. Fisher & his co-workers
  - Profound impact on agricultural science
- The industrial era, 1951 – late 1990s
  - Response surfaces, optimal design, robust design
  - Applications in the chemical & process industries
  - using SOE in Quality improvement in many industries (automotiv, electronics, and semiconductors,etc)

Modern era, after 1990

- **Computer software for construction and evaluation of designs has improved greatly with many new features and capability**

-

Formal education in statistical experimental design is becoming part of many engineering programs in universities

-

Applications of designed experiments have grown to all scientific area.

-

Randomized Controlled Clinical Trials widely used in drug discovery and public health.

-

Today e-commerce companies routinely conduct on-line experiments via website/emails.

Design of Scientific studies (Ch 15.1)

- **Experimental** studie: *establishment of **cause-and-effect** relationships between one or more explanatory factors (predictors, Xs) and a response variable (outcome Y)*
  -

    Ex: randomized clinical trials (RCTs):

    * commonly used to establish one treatment is better than the control/placebo; gold standard for drug approval by FDA.

  - **Observational** studies: *one can use to establish **association** between the explanatory factors and the response variable, but not causation.*

Comparative Experimental study

- In a *comparative experimental study*, **randomization** is employed to assign a set of treatmems to the experimental units, and the observed ourcomes among the treatment groups are compared to assess treatment effects. The **treatments** are defined by the levels of one or more explanatory factors, referred to as **experimental factors**. **Cause-and-effect** relationships between the experimental factors and the outcome or response variable can be established in an experimental study.

— Their difference: Randomization or not —

- An observational study : randomization of the treatment to experimental units does not occur.

SOE: Basic Concepts (Ch 15.2)

The structure of the experiment/study:

- The set of explanatory factors & treatments
-

The set of experimental units included.

```
+  The smallest experimental unit to which a treatment can be assigned, eg. a patient.
+  In a cluster randomized clinical trials, the units may be schools, hospitals, communities.
```

- **The rules and procedures by which the treatments are randomly assigned to the experimental units.**

```
+  Sample size per treatment or replication
+  Simple vs. blocked randomization.
```

- The outcome measurements that are made on the experimental units

Example 1 : Quick bread experiement

- **(p 678): A experiment was conducted to study the effect of baking temperature on the volume of a quick bread prepared from a package mix. Four oven remperatures-Iow. medium, high, and very high-were tested by randomly assigning each temperature to five package mixes.**

- **Q: What are the factor levels, sample size, experiment unit?**

Example 2 : A large clinical trial

- **The Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial**

- **10,251 participants with type 2 diabetes and cardiovascular disease or additional risk factors for cardiovascular disease was randomly assigned to either standard therapy (n=5,123) targeting HbA1c levels of 7.0%-7.9% or intensive therapy (n=5,128) targeting HbA1c $<$ 6.0%.**

- **Goal of this glycemia trial was to examine whether intensive therapy will reduce the rate of cardiovascular disease. events, compared to standard therapy**

- **Q: What are the factor levels, sample size, experiment unit?**

Example 3 : “Roadkill Experiment”

- **Studying which animals (snake vs turtle vs tarantula) are more likely to be run over by vehicles ?**

- **Any consideration to establish *cause-and-effect* ?**

Example 4 : Sample size and power

Overview of SOE (ch 15.3)

***We will illustrate with examples in the following weeks*:**

- Completely randomized design
- Factorial Experiments
- Randomized Block Designs
- Repeated Measures Designs
- Muti-Factor Experiments, etc

Software setup and homework

- Software setup (ppt slides)
- R demo for basic summary statistics
- Weekly Reading: Chapter 15.1-15.5
- Homework: Problem 15.9, 15.14 (due 8/27)