Introduction: Statistics and Engineering

STAT:2020
Probability and Statistics for Engineering and Physical Sciences

Week 1 - Lecture 1
Book Sections 1.1-1.2.4, 1.3: Introduction
Where do engineering and statistics meet?

**Biomedical engineering**

- Create an algorithm to diagnose eye disease based on photographs of the eye.
- Can I say that my algorithm does a good job at diagnosing? Does it do as well as a doctor who looks at the photograph?
- Collect photos from sample of patients, get diagnoses from both methods, compare to the truth. See if the data suggest the algorithm is reasonably accurate.
Where do engineering and statistics meet?

- **Civil engineering**
  - Which intersection should be improved first if we want to most efficiently reduce traffic accidents?
  - Are there certain characteristics (or combinations of characteristics) that make an intersection very dangerous? Turning lanes, traffic volume, etc. (study of 357 Chicago intersections below)

<table>
<thead>
<tr>
<th>Variable description</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work zone (1 if yes, 0 otherwise)</td>
<td>0.163</td>
<td>0.369</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Average annual traffic on major road (in thousands)</td>
<td>65.824</td>
<td>28.289</td>
<td>23.821</td>
<td>298.861</td>
</tr>
<tr>
<td>Snowy weather condition (1 if yes, 0 otherwise)</td>
<td>0.502</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Average annual traffic on minor road (in thousands)</td>
<td>37.185</td>
<td>13.964</td>
<td>4.901</td>
<td>99.750</td>
</tr>
<tr>
<td>Intersection without street lights in the evening (1 if yes, 0 otherwise)</td>
<td>0.574</td>
<td>0.495</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Weekday evening peak periods (5 p.m.–7 p.m.) (1 if yes, 0 otherwise)</td>
<td>0.799</td>
<td>0.401</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Defected pavement surface (1 if yes, 0 otherwise)</td>
<td>0.805</td>
<td>0.396</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Number of crashes</td>
<td>17.40</td>
<td>4.11</td>
<td>2</td>
<td>106</td>
</tr>
</tbody>
</table>
Where do engineering and statistics meet?

- **Industrial engineering**
  - In a work zone, does the type of lighting present affect a driver’s speed?
  - Collect data in a driving simulator under varying conditions (see NADS), fit a statistical model, determine if lighting plays a role in speed, after accounting for other variables that may affect speed (age, experience, etc)
At the heart of these questions is the idea of **variability**.

- Does the algorithm do better or worse under certain circumstances?

- The number of accidents at an intersection varies from year to year and across different intersections, but is that variability connected to certain things, such as traffic volume and number of lanes?

What are the potential **sources of variability** in number of accidents?

- There is a lot of **variability** in how fast people drive, but can that variability be explained by the lighting that is present? Or is there a more significant source of the variability in speed?
Chapter 1: INTRODUCTION

- An engineer is someone who solves problems of interest to society by application of scientific principles.
  - Listen
  - Communicate
  - Understand (science, mechanisms, language,...)
  - Be innovative (create/improve products)

- Scientific approach in engineering tends to be iterative (adjust, collect data, adjust, collect data,...). Often looking for an optimum.

- **Statistics**: collection, presentation, analysis, and use of data to make decisions, solve problems, and design products and processes.
Successive observations of a system do not produce the exact same results. Examples...

Can we better understand variability? Can we model variability?

How do we include it in our decision making process?

Statistics gives us a framework for describing variability, and we often consider many different sources of variability in a system.

Variability is a key concept in this course.
Chapter 1: INTRODUCTION

- **Visualizing variability**
  - Dot Diagram (~20 observations or less)
    Location and scatter are apparent

![Dot Diagram](image)

- Histogram (shows frequency of observed values)

![Histogram](image)
Chapter 1: INTRODUCTION

- **Visualizing variability**
  - Statistical software now has high capability for generating data-driven visualizations.
  
  Example: Stephen Curry, NBA player, FG% relative to the league average within each region of the court:

“BallR: Interactive NBA Shot Charts with R and Shiny” by Todd W. Schneider

Chapter 1: INTRODUCTION

- **Types of studies or experiments**
  - Retrospective study (in the past)
  - Prospective study (in the future)

- Observational study (hands-off)
  - Where you just observe, you don’t manipulate or change things directly.

- Designed experiment (manipulation by researcher)
  - Experimenter imposes changes, not just passive observer.
  - Randomization is used to establish cause and effect.
  - Plays an important role in manufacturing designs and development.
Statistical Inference

Statistical inference is when we infer something about the population from the information in a sample.

We usually want to say something about the population as a whole rather than the sample.

Data collection takes time, can be costly, uses resources, etc.
Mechanistic or Empirical Models

- Mechanistic model: \( \text{Current} = \text{voltage/} \text{resistance} \) or \( I = E / R \)

More realistic mechanistic model: \( I = E / R + \epsilon \)

- Empirical model (derived from data):

\[
\log(\text{HousePrice}) = \beta_0 + \beta_1(\text{squarefootage}) + \epsilon
\]
Chapter 1: INTRODUCTION

- **Probability Models**
  - A *probability model* is a mathematical representation of a random phenomenon (where outcomes are uncertain). It is a probability distribution related to the possible outcomes.

  - e.g. Normal Distribution \( f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \)

- These models help us quantify risk in the decision process (for decisions based on sampling)