Introduction to Statistics

Lulu Kang, MATH 100
What is Statistics

- wiki: Statistics is the study of the collection, organization, analysis, interpretation, and presentation of data.

- Statistics is *everywhere* in our life.
Examples: politics

- General Election: Romney v.s. Obama
- States Map
- Who's going where
Examples: Watch Sandy

- Sandy Path
Examples: Watch Sandy

- **Power Outage**

Power Outages during Hurricane Sandy

![Maps showing power outages during Hurricane Sandy](image-url)
Examples: When did Michael Jackson have his biggest hits?

Statistics is More Than Just About Data

• Statistics deals with all aspects of data, including planning of data collection.
• Design of Survey
• Design of Experiments.
Design of Survey

• How to construct effective surveys.

Here is an example of a double-barreled question:

<table>
<thead>
<tr>
<th>Bad Question: Double-barreled Question</th>
<th>Good Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>How have teachers and students at your school responded to the new 45-minute lunch period? ( ) Satisfied ( ) Unsatisfied</td>
<td>How have teachers at your school reacted to the new 45-minute lunch period? ( ) Satisfied ( ) Unsatisfied</td>
</tr>
<tr>
<td>How have students at your school reacted to the new 45-minute lunch period? ( ) Satisfied ( ) Unsatisfied</td>
<td>How have students at your school reacted to the new 45-minute lunch period? ( ) Satisfied ( ) Unsatisfied</td>
</tr>
</tbody>
</table>

Example 1 Balanced:

<table>
<thead>
<tr>
<th>Very Poor</th>
<th>Poor</th>
<th>Average</th>
<th>Good</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Example 2 Unbalanced:

<table>
<thead>
<tr>
<th>Poor</th>
<th>Average</th>
<th>Good</th>
<th>Very Good</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Design of Experiments

- What is the best recipe?
  - Oven Temperature
  - Sugar
  - Flour
  - Eggs

http://www.moresteam.com/toolbox/design-of-experiments.cfm#purposeExperimentation
Data Analysis

- There are so many data analysis techniques. Simplest one: linear regression.
Data Analysis

- Time Series Model: \( y \)-hotel sales v.s. months
Data Analysis

• Classification

**FIGURE 4.1.** The left plot shows some data from three classes, with linear decision boundaries found by linear discriminant analysis. The right plot shows quadratic decision boundaries. These were obtained by finding linear boundaries in the five-dimensional space $X_1, X_2, X_1X_2, X_1^2, X_2^2$. Linear inequalities in this space are quadratic inequalities in the original space.
Data Analysis

• Tree Model

FIGURE 9.5. The pruned tree for the spam example. The split variables are shown in blue on the branches, and the classification is shown in every node. The numbers under the terminal nodes indicate misclassification rates on the test data.
Data Analysis

• Graphical Model

Diagram:

- History of Smoking (HS)
- Chronic Bronchitis (CB)
- Lung Cancer (LC)
- Fatigue (F)
- Weight Loss (WL)
Data Visualization

• How to present data: more than just important!
• What's your Economic Outlook?
• The iPhone Economy
• Degree of Debt
• Facebook IPO
How Statistics can do to you?

- “I keep saying that the sexy job in the next 10 years will be statisticians,” said Hal Varian, chief economist at Google. “And I’m not kidding.” (http://www.nytimes.com/2009/08/06/technology/06stats.html)

- http://www.youtube.com/watch?v=D4FQsYTbLoI
To begin with

- MATH 474: probability and statistics
- MATH 476: Statistics
- MATH 484: Regression and Forecasting
- MATH 569: Statistical Learning
- ...