Big Data, Small Data: ... and My Dog Tito

Sanford Weisberg
sandy@umn.edu
z.umn.edu/clastat

University of Minnesota, Twin Cities
School of Statistics

July 10, 2015
How many pounds is Tito?
Give a 50% prediction interval.
Form groups of 3 or 4, and repeat these two questions.
On Your Own

- Goal may not be clear: what is a “pound”? 
- The “model”: Use what you know 
- Bias and precision 
- Prediction interval: How did you do it?
In a Group

- How did you get a consensus prediction?
  1. Use the mean
  2. Use a weighting scheme
  3. Use the median or minimum or maximum
  4. Use the bully method

For any of these, should the consensus prediction be better than the individual prediction?

- How did you get a consensus interval?
  1. Are group members correlated? Does this matter?
  2. Average individual intervals
  3. Try to include between subject variability
  4. Some other way

Should these work?
Statistics is all about collecting, combining, and interpreting numerical information.
In 1974, psychologist Voevodsky ran an experiment:

- 343 S. F. taxis fitted with center break light, and 160 as control

- Drivers assigned to taxis at random, regardless of driver preference.
- After 10 months, taxis with center light had about 60% fewer accidents, fewer injuries, lower repair costs.
- Experiment repeated on large scale by NHTSA with same results, and center break lights were mandated in 1986 for cars and 1994 for light trucks.
Long-Term Evaluation of Center Brake Lights

<table>
<thead>
<tr>
<th>Accident reduction, experimental period</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accident reduction, follow up period</td>
<td>5%</td>
</tr>
</tbody>
</table>

What happened?
Another (likely) example: The Netflix challenge
http://en.wikipedia.org/wiki/Netflix_Prize

In 2009, Netflix offered a $1 million prize for improving prediction of user ratings of movies by 10%.

Data: 100 million ratings of 18 thousand movies by 480 thousand customers.

Winning Methodology: Weighted average of many algorithms. Tested (by Netflix) on data withheld from the competitors.

Will this method continue to work well now and in the future?
Predict average quality rating of professor, 1-5 scale, to be predicted, using:

- Professor’s gender
- **Number of years** this professor is included in the website
- **Discipline**: humanities, social science, pre-professional, STEM
- Average **easiness** of courses, 1-5 scale
- Average **rater interest** in subject matter, 1-5 scale
- Physical **attractiveness** scale, yes/no

Data consists of average ratings of 366 professors at a large campus in the Midwest.
World-class faculty

Degrees awarded: about 100 undergraduate, 20 masters, and 7 PhD per year

New! Bachelor of Arts in the Practice of Statistics

New! Bachelor of Science in Statistical Science

Minor in Statistics
Bachelor of Arts in the Practice of Statistics

- Designed as (1) a terminal degree; (2) a second degree; (3) preparation for medical school, law school, business school, graduate school in other than statistics

- Available only at the University of Minnesota

- Minimal prereqs: 1 year of calculus, 1 semester C or C++, 1 semester of “Stat 1”

- New! majors-only core curriculum in applied statistics equivalent to most MS programs; new internship program; stress on communication

- Opportunity for including quantitative courses from areas like genetics, geography, psychology, and others in the major
Prereqs include LOTUS of mathematics, at least equivalent to a math minor

Share the majors-only applied statistics core and internship program with the BA

Elective courses will generally be even more mathematics and computer science

Designed for students planning graduate work in a statistical science: statistics, data science, actuarial science, psychometrics, biostatistics, decision sciences, others
The 10 Best Jobs of 2015


1. Actuary – $94,209
2. Audiologist – $71,133
3. Mathematician – $102,182
4. Statistician – $79,191
5. Biomedical Engineer – $89,165
6. Data Scientist – $124,149
7. Dental Hygienist – $71,102
8. Software Engineer – $93,113
9. Occupational Therapist – $77,114
10. Computer Systems Analyst – $81,150
“I keep saying the sexy job in the next ten years will be statisticians. People think I’m joking, but who would’ve guessed that computer engineers would’ve been the sexy job of the 1990s?”

—Hal Varian, Google Chief Economist, *The McKinsey Quarterly*

Great source for information on being a statistician: 

amstat.org/careers