

**NUMBERS IN EVERYDAY LIFE  
FURTHER EXAMPLES AND WRAP-UP**

Union College Academy for Lifelong Learning  
Class 5

Gerry Hahn  
Jane Oppenlander  
May 1, 2008

**TALK OUTLINE**

- Quantifying the subjective: College rankings
- Testing in schools
- More on data mining
  - Wal-Mart’s data warehouse
  - Data mining and personal privacy
  - Coming Attraction: Pollution in the Lake Champlain Watershed
- “One in a million chance” event
- More on good and bad numbers (in brief)
  - Sports applications
  - Good and bad graphics
  - Misleading numbers: Some further scenarios
  - Number studies that further knowledge
  - Some good reading and surfing
- Course Take-Aways



**QUANTIFYING THE SUBJECTIVE  
—COLLEGE RANKINGS**



- U.S. News and World Report provides yearly college rankings
- Union ranks 40<sup>th</sup> among 266 Liberal Arts colleges
- RPI ranks 44<sup>th</sup> among 262 universities
- Similar issues arise in ranking
  - Other service providers, e.g., hospitals
  - Consumer products
  - Movies
  - Employees

WHAT IS THE BASIS OF THE RANKINGS?

**CRITERIA**



- Peer assessment (from reviews by college presidents, provosts, deans of admission): 25%
- Various numbers (based on responses to questionnaires sent to colleges):
  - Graduation and retention rate: 20%
  - Student selectivity (SAT/ACT scores, high school standing, acceptance rates): 15%
  - Faculty resources (Class size, compensation, top degree, %full-time, student/faculty ratio): 20%
  - Financial resources: 10%
  - Alumni giving: 5%
  - Graduate performance: 5%

**ISSUES AND CONCERNS**  
(See Best, 2004)



- What do we mean by “best?”
- Are the right criteria and weights being used?
- Emphasis on what can be measured
- Incentive to colleges to “game” the system
- Also may not reflect *your* criteria

**BASIC CONCEPTS**



- Einstein: Not everything that can be measured is important, and not everything that is important can be measured
- Numerical rankings of
  - Service providers is difficult
  - Products sometimes less difficult
- Need to ask: How were rankings developed?
- Relate to *your* value system—possibly via user-supplied weighting system

### TESTING IN SCHOOLS



- **No Child Left Behind (NCLB) Act (2001)**: Calls for every child on board by 2014 ("Each school must improve each year until ALL students meet...standards.")
- Testing mandated to monitor progress
- Yearly test results used to
  - Evaluate schools (and school districts) : In good standing, need for improvement, etc.
  - Identify weakness areas
  - Help determine remedial funding
- Different viewpoints
  - Former Secretary of Education Paige: "Anyone who opposes annual testing is apologist for a broken system of education"
  - Joel Best (author of *Flavor of the Month*, 2006): The current big fad.
- Controversy involves both numbers (our focus) and other issues (e.g., teaching to test, inadequacy of test questions, inhibition of creativity, emphasis on selected subjects)

### NITTY-GRITTIES OF METHODOLOGY

N.Y. State, 2007-8: English Language Arts (ELA) and Mathematics



- **All students tested yearly in grades 3 through 8 and beyond**
- **Test result quantified as Level 1, 2, 3 or 4 (best)**
- **School evaluated on each group with 30 or more**
  - All students
  - 6 ethnic groups
  - 3 other groups (disabilities, limited English, disadvantaged)

### MORE NITTY-GRITTIES

- **Group Performance Index (PI)** calculated for each group:
  - %Kids at Level 2 + 2 x (% Kids at Levels 3 and 4)
  - Example for group of 100 students with 10, 30, 40 and 20 kids at Levels 1, 2, 3 and 4, respectively:  
 $PI = 30 + 2 \times (40 + 20) = 150$
  - Maximum Possible: 200 (ALL students at Level 3 or 4)
- **Annual Measurable Objective (AMO)** set in NY State as
  - 2006/7 ELA: 122
  - 2007/8 ELA: 133
  - 2013/14: ELA (and Mathematics): 200 (perfection!)
- **Effective AMO**: 95% lower confidence bound based on group size:
  - 2007/8 for ELA group of 30-34 students: 116 (versus 133)
  - 2007/8 for ELA group of 120-149 students: 124 (versus 133)
- **PI must exceed effective AMO for ALL groups** (under review, NY Times, March 19, 2008) with 30 or more students

### SOME CONSEQUENCES



- **Goal: Raise Group Performance Index (PI)** for each group:
  - $PI = \%Kids \text{ at Level } 2 + 2 \times (\% \text{ Kids at Levels } 3 \text{ and } 4)$
- **Improve scores by**
  - Raising Level 1 kids to Level 2 and then Level 3
  - Raising Level 2 kids to Level 3
- **No gain** from raising Level 3 kids to Level 4!

### SOME LIMITATIONS OF METHOD



- **No incentive to advance Level 3 kids to Level 4**
- **Tool to compare schools, districts, etc: Does not consider socio-economic background of students, parent support, etc.**
- **Tool to compare progress of schools, etc: Might be impacted by change in demographics**
- **School dropout rates**
  - Not part of PI calculation (might encourage Levels 1&2 dropouts!)
  - Inconsistent reporting by different States (NY Times, March 27, 2008); to be corrected by 2012-13 (NY Times, April 28, 2008)
- **Requires perfection by 2014!**

### AN IMPROVEMENT: ADDED-VALUE ASSESSMENT



- **Based on change in individual student yearly test scores: Neutralizes impact of many other factors**
- **Uses complex statistical model (Henderson mixed-model equations) via SAS computer program (EVAAS). This**
  - Allows use of all available info on each student (even if some data are missing)
  - Provides claimed "statistically unbiased estimates of the influences of districts, schools, and teachers on the rate of academic progress"
- **Status:**
  - Developed and implemented in 1992 in Tennessee
  - Mandated in Pennsylvania and Ohio (and various school districts)
  - Arkansas and Minnesota getting on board
  - Piloted in other States including New York (NY Times, Jan 21, 2008)
- **Not perfect and controversial (also complex and expensive)—but likely better than current approach**

## BASIC CONCEPTS

- **Need for quantifying achievement and progress**
- **Limitations of current NCLB assessments**
  - Inadequacies of numbers
  - Beyond the numbers
- **The likely best approach**
  - Seek most meaningful numbers possible
  - Supplement these with systems knowledge and apply sensibly

### “What Wal-Mart Knows About Consumer Habits”

The New York Times, Nov. 14, 2004



- In 2004 Hurricane Frances was moving rapidly across the Caribbean aiming for a direct hit on the Atlantic Coast.



**While in Bentonville, Arkansas . . .**

- Wal-Mart was preparing for the storm by combing through their massive database to see what customers bought before a previous hurricane – Charley.
- They expected a run on flashlights and batteries.



### What Wal-Mart learned by mining its sales data?

- **Top selling item during Hurricane Charley**



Beer



- **An item that sold at 7 times its normal rate before Hurricane Charley**

Strawberry Pop-tarts



### How Wal-Mart helped its bottom line

- Trucks headed off to Wal-Mart stores in the path of Hurricane Frances bearing flashlights, batteries, beer, and strawberry Pop-tarts.



- According to Wal-Mart, the beer and Pop-tarts sold quickly.
- “Such knowledge, Wal-Mart has learned, is not only power. It is profit too” - The New York Times

## DATA MINING REVIEW

- **Automated** process of collecting and **analyzing large volumes** of data to find **hidden patterns**.



- Where is all the data kept?

### Databases Used by Retailers

- **Transactional Database**
  - Supports day-to-day business operations
  - Designed to be efficient for relatively simple, repetitive transactions (taking an order over the telephone, pay a vendor invoice)
  - Data is very dynamic
  - May be a separate database for each store, region, or district



### Databases Used by Retailers

- **Data Warehouse**
  - Supports business decision-making
  - Designed for data mining and complicated data queries. (How many 18 oz. boxes of Wheaties were sold in the Albany area stores last week?)
  - Data is updated periodically
  - Typically one warehouse for the whole enterprise
  - Contains transaction data and more (facility information, customer demographics, etc.)




### How big is Walmart's data warehouse?

- **Biggest data warehouse in the world**
- **Records every item purchased at check out**
- **~800 million transactions/day for 4000 retail stores**
- **Over 460 terabytes of data (1 terabyte is a trillion bytes)**
- **Get customer information from cashing payroll checks (need SSN), Sam's Club Membership information, check cashing (need drivers license)**



### What does Wal-Mart do with the data?

- **Manages suppliers and inventory**
  - Allows suppliers direct, but limited access to Wal-Mart's data
- **Looks for customer buying patterns**
- **Evaluates store, employee, and regional performance and facility layouts**

Data mining → Business Intelligence → \$\$\$

### How Important is the data warehouse to Wal-Mart?

- **Basis for the legendary Saturday sales meetings and supplier negotiations**
- **Companies that sell their computer hardware and software to Wal-Mart must sign a non-disclosure agreement**
- **Spent 4 billion dollars for data warehouse**
- **Will not reveal any information about their data warehouse or data mining methods**
- **Maintain their own IT staff**
- **Funds university programs in data warehousing technology**

### The Wal-Mart Data Center

- **Built on bedrock**
- **Unmarked**
- **One story**
- **15 miles from HQ**



- **Back-up generators, fuel, sleeping quarters**
- **Fences have razor wire on top**
- **Earth built up against walls**

### Bottom Line

- **For Wal-Mart**
  - PROFIT
  - Increased markets
- **For consumers**
  - Everyday low prices
- **For civil libertarians**
  - Concern about concentration of so much information in one company.
  - Potential for abuse of personal data





### Data Mining and Personal Privacy

- Do large data bases and data warehouses “owned” by businesses or the government pose risks for violations of personal privacy?
- What about the ability to obtain and integrate data from multiple databases?
- Do the benefits of data mining outweigh the risks? Can abuses be controlled?



Are you planning to order pizza for dinner tonight?

### Cause for Concern?

- Wal-Mart’s chief compliance officer and NYC mayor Michael Bloomberg announce Wal-Mart’s new tougher rules for firearms sales.
- Wal-Mart will videotape all firearms sales.
  - Create an alert system to record when a gun sold at Wal-Mart is later used in a crime. Prevent additional sales to buyers of those guns.
  - Videos and records saved in case law enforcement wishes to view them as part of an investigation.
  - Expanded background checks of employees who handle guns.



Wal-Mart has yet to establish a store in NYC

### Homeland Security and the Search for Terrorists

- Various government programs to provide tools to detect, classify, and identify potential foreign terrorists.
- MATRIX database contained 3.9 billion public records collected from thousands of sources.
  - FAA pilot license and aircraft ownership records
  - Vessels registered with the Coast Guard
  - Sexual offender lists
  - Bankruptcy filings
  - State-issued professional license records
  - Motor vehicle and drivers’ license information
  - Department of Corrections information



These programs have been controversial

### Some of the Issues

- Identification of **known** terrorists vs. identification of future terrorists by **profiling**
- Use of data for purposes other than that for which it was originally intended without consent of the individual
- Quality and accuracy of the mined data
- Falsely identifying individuals as terrorists



### Cause for Concern?

- In 2005 and 2006 the FBI mined sales data from San Francisco area grocery stores looking for Iranian terror cells.
  - Believed a spike in sales of Middle Eastern food would lead to them to terrorists.
  - It is not clear whether the groceries or credit card companies gave up the data voluntarily or through the Patriot Act.
  - Program was cancelled by the head of the FBI’s criminal investigations division, Michael A. Mason.

Mason left the FBI to come head of security for Verizon

### “Price Chopper alerts its customers to recall using data from discount cards”

Albany Times Union

- Price Chopper with the help of southern California company Smart Reply called 18,000 households likely to have purchased Samuel Adams beer that was recalled for potentially containing shards of glass.
- System was used:
  - In February when Ice Hot Heat Therapy items were recalled.
  - In March when Stonyfield Farms blueberry yogurt was recalled for containing fragments of glass or plastic.



**How do retailers get your personal information?**

- Rewards or loyalty cards.
- Check cashing
- Retailer sponsored credit cards
- Warehouse club memberships
- Rebates
- Web sites



**Some Web Statistics**

- A study by Georgetown University sampled 361 commercial web sites and found:
  - 92.8% collected personal identifying data
  - 56.8% collected demographic information
- There is virtually no cost to collect information over the Internet and little regulation.
- When accessing web sites look for privacy information:
  - Notice
  - Choice
  - Access
  - Security
  - Contact Information

**Coming Attractions**



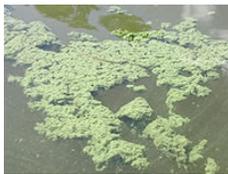
**Pollution in the Lake Champlain Watershed**



- 120 miles long
- 3 jurisdictions – New York, Vermont, Quebec
- Up to 400 feet deep

Health of the lake is vital to the region's economy

**Problems in the Lake Champlain Watershed**



- Excessive phosphorus pollution leads to algae blooms which chokes lake habitat



- Invasive species



**A Data Mining Approach**

- \$6.7 million National Science Foundation grant for Complex Systems Modeling for Environmental Problem Solving.
- No new data will be collected.
- Build a database of all available information, e.g., fish habitat studies, storm water runoff analyses.
- Use data mining to reveal hidden patterns that emerge from the complex interaction between water, organisms, pollution, weather, . . .
- Plan to study from the molecular level to the global scale.



### Complex vs. Complicated

#### Complicated

- A watch has many parts. Remove 1 gear and predictably it stops working.

#### Complex

- Pollute one river. Can not predict when or where an algae bloom may occur.

### BASIC CONCEPTS



- Data warehouses enable effective and efficient data mining.
- Data mining is inherently observational. "Past performance does not guarantee future returns."



- Be wary of those who seek your data.

### "ONE IN A MILLION CHANCE" EVENT



Nationally reported news item Feb 7, 2008:  
 "In Syracuse Democratic primary Hillary Rodham Clinton and Barack Obama each got 6,001 votes ....The odds of that happening are less than one in a million, said Syracuse University mathematics professor Hyune-Ju Kim."

CAN THIS BE CORRECT?

### OUR EVALUATION



- Assessment depends on underlying assumptions
- Consider analogy of 12,002 coin tosses
  - Under highly conservative assumption that each of 12,003 outcomes is equally likely: Chances are 1 in 12,003 of getting exactly 6001 heads in 12,002 tosses
  - If we assume "fair" tossing of "fair" coin (i.e. probability of heads in each toss is 0.5): Chances are 1 in 137 of getting exactly 6001 heads in 12,002 tosses
- So how did professor come up with 1 in a million? (per email exchange)
  - She assumed that Syracuse voters were random sample of NY State voters (59% for Clinton; 41% for Obama)
  - She addressed question "what are chances of getting 6001 heads in 12,002 tosses of a coin for which probability of getting heads is 0.59?"
  - She communicated her assumptions to reporter (who failed to mention it in report)

### SOME FURTHER CONSIDERATIONS

- Chance of even split in Syracuse is still small
- However, there are an estimated 2,000 localities in U.S.
- Chances of even split *somewhere* in the U.S. are good

### BASIC CONCEPTS



- Seek precise definition of what numbers claim is saying and underlying assumptions
- Try some simple ideas to make your own assessment
- Don't assume it is right because it appears in newspaper
- Impressive and surprising results are what make news (Best, 2004)

**MORE ON GOOD AND BAD NUMBERS  
SPORTS APPLICATIONS**



- **Baseball**
  - A numbers person's dream
  - Some reflect variables beyond a player's ability
  - No popular "super-statistic"
- **Numbers used in**
  - **Game strategy**
    - Basketball: Player-on-floor decisions (e.g., Advanced Scout)
    - Baseball: The deliberate walk
    - Football: Go for 4th down
  - **Front office strategy** Building the Oakland A's on a shoestring (per M.Lewis: Moneyball)
  - **Fan speculation**

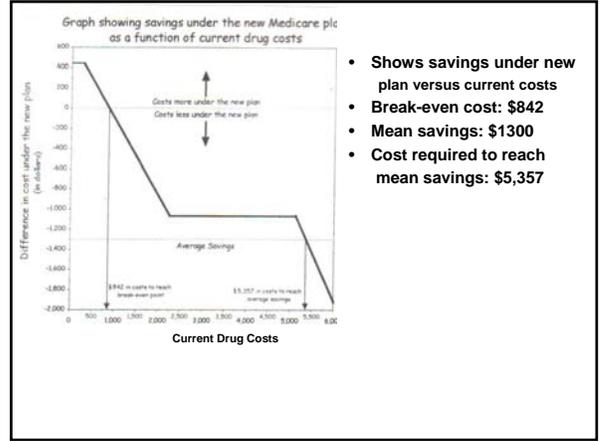
**WILL 0.300 HITTER WITH 0.350 BATTING AVERAGE IN FIRST 100 AT BATS END SEASON WITH 0.350 AVERAGE?**



- Chances that 0.300 hitter is hitting 0.350 (or better) after 100 at bats are about 12%
- But chances 0.300 hitter will hit 0.350 in remaining 550 season at bats are less than one in 200
- Calculation assumes that 0.300 hitter remains 0.300 hitter

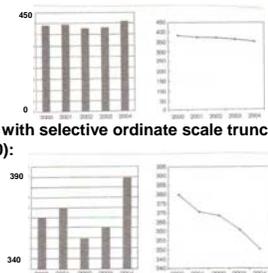
**GOOD AND BAD GRAPHICS**

- A good picture is worth a thousand statistics (Wayne Nelson)
- Example: Another look at Medicare drug plan (see Wainer, *Chance*, Spring 2006)
  - Bush: "On the average the folks who sign up ...are going to save \$1,300 a year"
  - Average was taken as mean (not median)
  - Graphics often can convey story simply



**GRAPHICS CAN ALSO MISLEAD**

- One example: Scale truncation
- Example from Best (2004)
  - Bar graph and line graph starting at 0:
  - Same data with selective ordinate scale truncation (starting around 340):



**MISLEADING NUMBERS: SOME FURTHER SCENARIOS**  
(From Best, 2001 and 2004)



- **Inadequate measurement and/or definition**
  - Medical errors kill between 40,000 and 98,000 U.S. hospital patients each year
  - Estimating number of Muslims or Jews in the U.S.
- **Changes in record keeping and definitions over time**
  - Incidents of child abuse
  - Incidence of autism
  - Incidence of homelessness
- **Differences in record keeping or situations over different places**
  - Incidence of child abuse, autism and homelessness in different States
  - Number lawyers in U.S. versus Japan
  - US high school students score worse than their counterparts in other countries
- **Differences in what is reported: "Data cherry-picking"**
  - Change in personal income (adjusted for inflation) from 1959 to 1999
    - Per capita income increased by 150% (and increased by 71% since 1974)
    - Hourly earnings increased by 17% (and decreased by 5% since 1974)

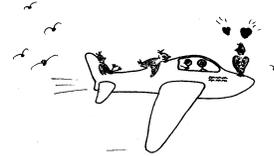
THERE ARE (AT LEAST) TWO STATISTICS TO EVERY STORY

**NUMBER STUDIES THAT FURTHER KNOWLEDGE**

- **Historical issues:**
  - Authorship of disputed Federalist papers: Hamilton versus Madison
  - Evaluating archeological finds: Does 1980 Jerusalem burial tomb find contain ossuaries (limestone coffins) linked to New Testament figures?
- **Current issues (mostly from Chance Magazine, 2006)**
  - Early detection of bioterrorism via statistical process control
  - Evaluation of impact of use of automobile airbags
  - Assessment of discrimination by gender, race, age, etc.
    - Salaries paid
    - Racial profiling (in driver stops and searches)
  - Evidence of global warming caused by human activity
- **Note: Such issues often require complex methods and may lead to controversial findings**

**NUMBER STUDIES THAT FURTHER KNOWLEDGE**

- **And some personal experiences**
  - Which brand food do dogs prefer?
  - Do birds mistake jet engine noise for mating calls?



**SOME GOOD READING AND SURFING**

- **General:** Best (2001), Best (2004), Peck et al (2006) and Utts (1999)
- **Misuse of Statistics:** Hooke (1983), Huff (1954)
- **Related Topics:** Levitt & Dubner (2005), Paulos (1991), Lewis (2003)
- **Getting the Picture:** Gonick and Smith (1993)
- **Magazine:** Chance
- **Web sites**
  - Chance News (Dartmouth)
  - Carl Bialik (Wall Street Journal)
- **Class 1 handout**
  - Gives summary comments
  - Lists others



**THE BLACK SWAN: COMMENTS**



- Nassim Nicholas Taleb: *The Black Swan*, Random House, 2007
- **Basic Thesis:** We need focus on the unusual and the critical (instead of the common and mundane)
- **Examples**
  - 9/11/2001
  - Hurricane Katrina
  - Black Swans
- **Some consequences**
  - The future is unpredictable
  - Forget about the bell-shaped curve
- **Commentary**
  - Some good points
  - But tends to throw out the baby with the bathwater

**FREAKONOMICS: COMMENTS**

- **Steven D. Levitt and Stephen J. Dubner:** *Freakonomics*, HarperCollins, 2005
- **Basic Thesis:** Thoughtful use of numbers can help unravel many of life's mysteries
- **Examples**
  - Identifying cheaters
  - Explaining the drop in crime rates
  - "Perfect" parenting
- **Comment:** Interesting (and thought-provoking) reading

**COURSE TAKE-AWAYS**



- Numbers are an essential and highly valuable element of numerous human endeavors—you can't escape them
- **Always ask**
  - Who is taking/reporting the numbers?
  - How were they obtained?
  - Have they been peer-reviewed?
  - What are the underlying assumptions?
- **Be wary of**
  - Advocates' numbers
  - Cherry-picking
  - Before and after comparisons
- Remember news media seek newsy/surprising numbers
- Appreciate limitations of observational studies and differentiate correlation from cause and effect
- Gold standard is controlled (randomized) experimentation—but often not attainable
- Recognize uncertainty: Nothing is certain, but death and taxes (Franklin)
- Let numbers help you gain understanding—not intimidate you!

**CLASS MOTTO:** Numbers are highly useful, but can be readily abused—handle with care!