Statistics 36-203: Sampling, Surveys and Society

Brian W. Junker
132E Baker Hall
8-8874
brian@stat.cmu.edu
http://www.stat.cmu.edu/~brian/303
http://www.cmu.edu/blackboard

Today

- Introductions
- Syllabus
- About the Class
- Course Outline
  - Team Project
- Readings
- Some History of Survey Sampling
- Examples

Introductions

- Instructor:

  Brian W Junker

  132E Baker Hall

  brian@stat.cmu.edu

- TA:

  Jionglin Wu

  FMS Building

  jionglin@stat.cmu.edu

- Web

  http://www.stat.cmu.edu/~brian/303

  http://www.cmu.edu/blackboard

About the Class

In this course we address questions such as:

- What methodology underlies sample surveys?
- How are surveys used and by whom?
  - government
  - business
  - media
  - universities (e.g., for management)
  - researchers
- What should you look for when you read about a survey?

Course Outline

1. Basic Concepts of Survey Design

- Gathering data
  - questionnaire design
  - modes of interviewing
- Dealing with those who don't respond

Course Outline (continued)

2. Course Projects

- Design, implement, analyze, and report on your own survey.
- Work as part of a survey team.
- Project work extends throughout the course.

Course Outline (continued)

- Team project assignment
  - On-campus vs. off-campus projects
- Getting started by forming a 4-5 person team
  - Form by Friday noon and plan meeting
  - Email Team Members to me: brian@stat.cmu.edu
  - Residual students will be assigned to groups by me

Course Outline (continued)

3. Basic Ideas of Sampling From a Finite Population

- using tools from 36-201/202 to understand the properties of samples
- idea of urn models
- stratification and clustering
- learning through computer simulation exercises

Course Outline (continued)

4. Case Studies: Examples
   - Current Population Survey
   - “Sex in America”
   - J.D. Powers Automobile Satisfaction Survey
   - 2010 US Census – Not a survey but many survey ideas are used!
   - Intersurvey -- web-based panel
- Sample designs, questionnaires, measurement problems, response rates, statistical inferences.

Syllabus

- Text
  - ***Required***: Groves et al: *Survey Methodology*
  - Other Useful Texts [not required for this class]:
    - Kalton: *Introduction to Survey Sampling*
    - Converse & Presser: *Survey Questions: Handcrafting the Standardized Questionnaire*
- Additional Readings [I'll provide]
  - Lohr: *Sampling: Design and Analysis*
  - Handouts
- Computing
  - MINITAB or R.
  - Other software (SAS, Excel, etc.) you are on your own
  - Possibly a special survey simulation program.

Syllabus (continued)

- Review Quiz on Thursday
- Grading
  - HW, quizzes, other assignments 20%
  - Two Exams (Feb 17, Apr 12) 30%
  - Citizenship/Discussion/Attendance 10%
  - Project Phase I 5%
  - Project Phase II 5%
  - Project Phase III (report & presentation) 30%

Readings

- Updated reading assignments will be maintained on http://www.stat.cmu.edu/~brian/303
- Begin by reading Groves, Chapters 1 & 2

Brief History

- Governments have always collected data on the state
  - Early *official statistics* date at least to 2200 BC China
  - *Statistics*, ca.1770: “affairs of state”, borrowed in 1829 to mean “numerical data collected and classified” (www.etymonline.com)
  - Used for taxation and other official record-keeping
  - *Census* (enumeration of every person) was the most common careful method
  - US Constitution: Decentennial census to reapportion House of Representatives

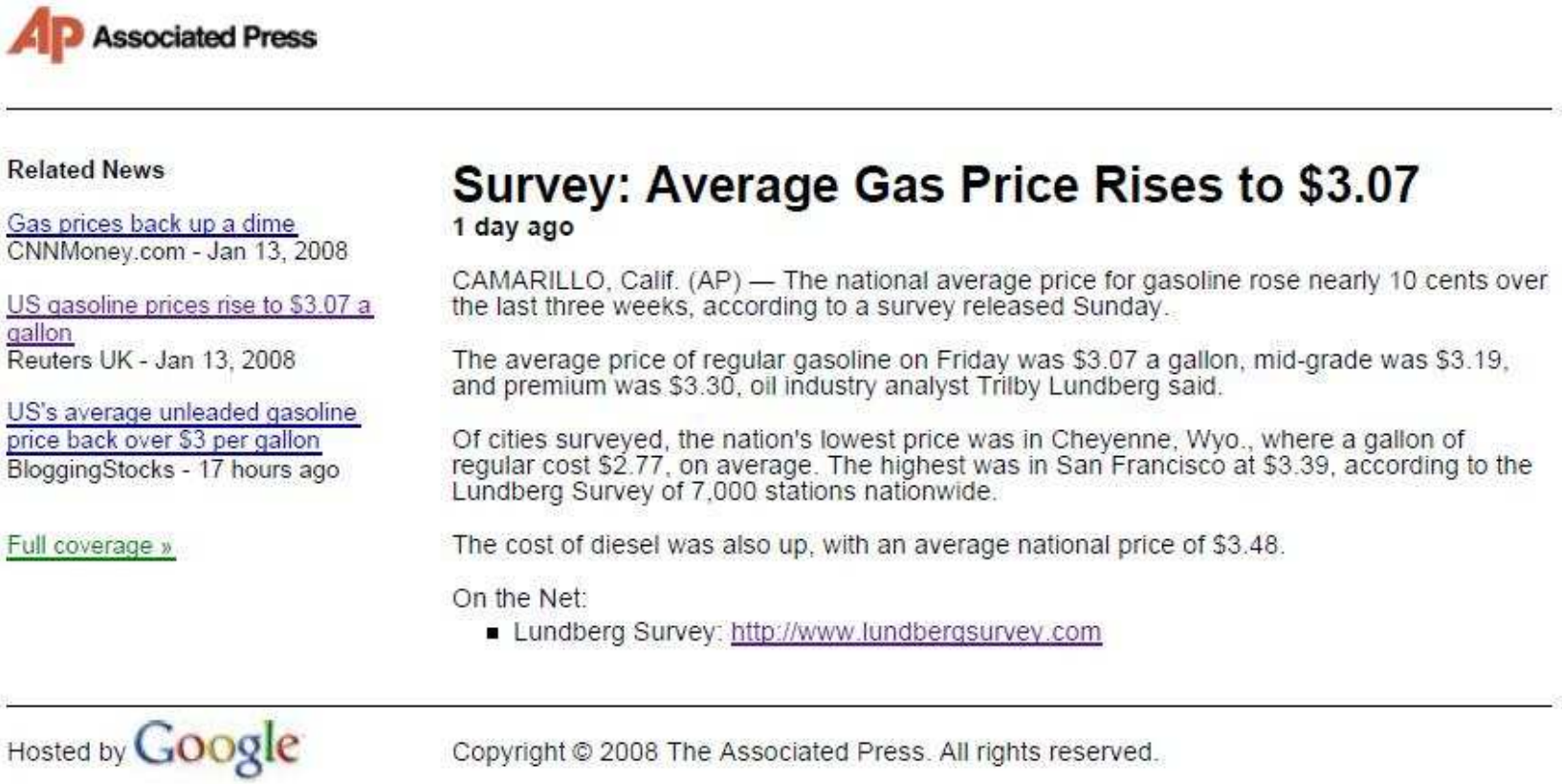
Brief History (continued)

- Emile Durkheim (1890's) used *secondary analysis* of official statistics to learn about social phenomena (division of labor, crime & punishment, education, religion, suicide)
- Charles Booth: *Life and Labor of the People of London (1889-1903)* was first "modern" survey (but still census was the goal)
- Early 20th Century: Serious debate about whether a *sample* could be as reliable as a *census*
- 1930's: journalists and market researchers began using survey methods to get a systematic view of "the man on the street"

Survey Research Today

- 65,122 hits on Google, e.g.:
  - Americans, Amid Recession, Dislike Their Jobs
  - Survey: Chile Consumer Price Index Seen Falling 0.3% In Dec
  - A University of Illinois at Chicago research group will perform a survey of River Forest residents to see what they think about police service
- US Government Monitoring National and Social Policy Effects
  - *The National Crime Victimization Survey*
  - *The National Assessment of Educational Progress*
  - *The Current Population Survey*
  - *2010 Decennial Census*
- Marketing Research (*Quirks* magazine)
  - Searching for representative Black households
  - How not to market to Asian-American Youth
  - Comparing online and phone research
  - Keep undesirable respondents out of your online survey
- Standard method in education, psychology and the social sciences

Example: Gas Prices



Associated Press

Related News

Gas prices back up a dime
CNNMoney.com - Jan 13, 2008

US gasoline prices rise to $3.07 a gallon
Reuters UK - Jan 13, 2008

US's average unleaded gasoline price back over $3 per gallon
BloggingStocks - 17 hours ago

Full coverage »

Survey: Average Gas Price Rises to $3.07

1 day ago

CAMARILLO, Calif. (AP) — The national average price for gasoline rose nearly 10 cents over the last three weeks, according to a survey released Sunday.

The average price of regular gasoline on Friday was $3.07 a gallon, mid-grade was $3.19, and premium was $3.30, oil industry analyst Trilby Lundberg said.

Of cities surveyed, the nation's lowest price was in Cheyenne, Wyo., where a gallon of regular cost $2.77, on average. The highest was in San Francisco at $3.39, according to the Lundberg Survey of 7,000 stations nationwide.

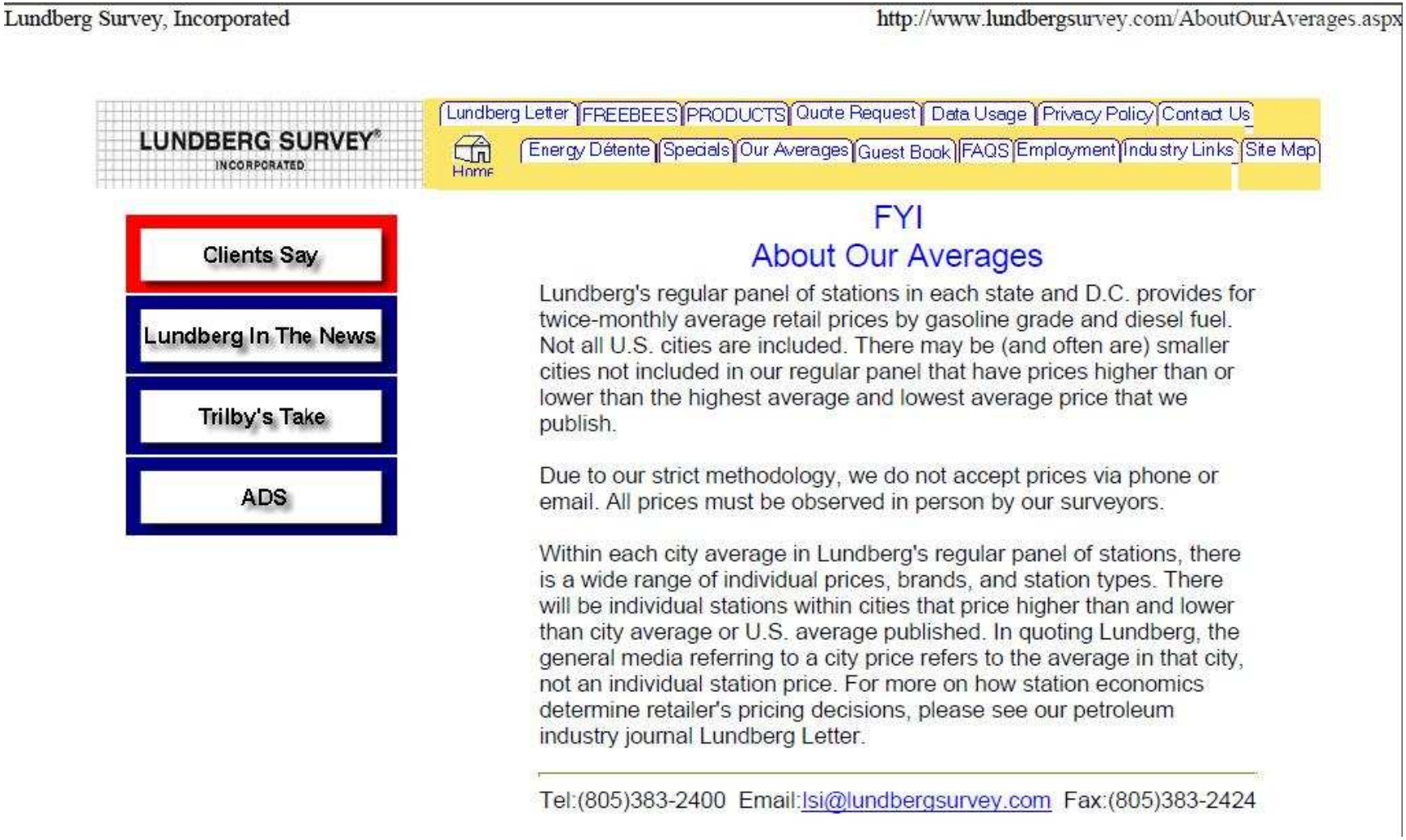
The cost of diesel was also up, with an average national price of $3.48.

On the Net:

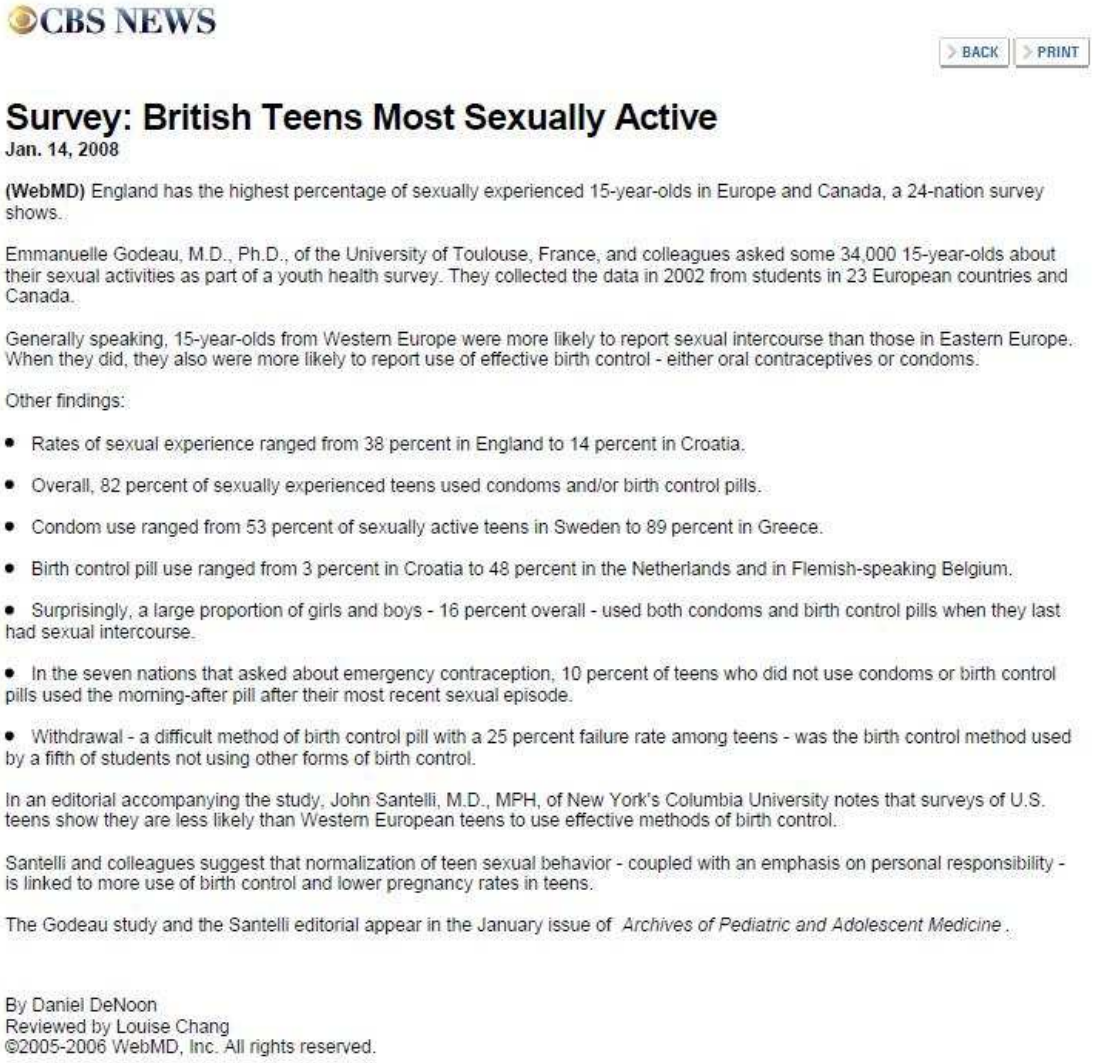
- Lundberg Survey: http://www.lundbergsurvey.com

Hosted by Google

Example: Gas Prices



Example: Health Behaviors of School-Age Children



CBS NEWS

BACK PRINT

Survey: British Teens Most Sexually Active

Jan. 14, 2008

(WebMD) England has the highest percentage of sexually experienced 15-year-olds in Europe and Canada, a 24-nation survey shows.

Emmanuelle Godeau, M.D., Ph.D., of the University of Toulouse, France, and colleagues asked some 34,000 15-year-olds about their sexual activities as part of a youth health survey. They collected the data in 2002 from students in 23 European countries and Canada.

Generally speaking, 15-year-olds from Western Europe were more likely to report sexual intercourse than those in Eastern Europe. When they did, they also were more likely to report use of effective birth control - either oral contraceptives or condoms.

Other findings:

- Rates of sexual experience ranged from 38 percent in England to 14 percent in Croatia.
- Overall, 82 percent of sexually experienced teens used condoms and/or birth control pills.
- Condom use ranged from 53 percent of sexually active teens in Sweden to 89 percent in Greece.
- Birth control pill use ranged from 3 percent in Croatia to 48 percent in the Netherlands and in Flemish-speaking Belgium.
- Surprisingly, a large proportion of girls and boys - 16 percent overall - used both condoms and birth control pills when they last had sexual intercourse.
- In the seven nations that asked about emergency contraception, 10 percent of teens who did not use condoms or birth control pills used the morning-after pill after their most recent sexual episode.
- Withdrawal - a difficult method of birth control pill with a 25 percent failure rate among teens - was the birth control method used by a fifth of students not using other forms of birth control.

In an editorial accompanying the study, John Santelli, M.D., MPH, of New York's Columbia University notes that surveys of U.S. teens show they are less likely than Western European teens to use effective methods of birth control.

Santelli and colleagues suggest that normalization of teen sexual behavior - coupled with an emphasis on personal responsibility - is linked to more use of birth control and lower pregnancy rates in teens.

The Godeau study and the Santelli editorial appear in the January issue of *Archives of Pediatric and Adolescent Medicine* .

By Daniel DeNoon
Reviewed by Louise Chang

Example: HBSC

The 2001/2002 HBSC survey

HBSC surveys are carried out at four-year intervals. Their findings are used to inform and influence health promotion and health education policy and practice for young people at the national and international levels, and to advance scientific knowledge. The 2001/2002 survey, on which this report is based, is the sixth in the series and the most recent. It was conducted successfully in 35 countries and regions (Table 1.1).

The data are collected in all participating countries and regions through school-based surveys, using the international research protocol. The survey instrument is a standard questionnaire developed by the international research network. The target population of the study comprises young people attending school, aged 11, 13 and 15 years. These three age groups represent the onset of adolescence, the time when young people face the challenges of physical and emotional change; and the middle years, when important life and career decisions are beginning to be made.

The questionnaire consists of a set of mandatory items that each country or region must use to facilitate the collection of a common set of data. These are the data presented in this international report. In addition, optional questionnaire items cover specific topics. Data from these will be reported in future publications.

Annex 1 describes the survey methods in detail. A brief summary is given here.

Sampling and data collection

The 2001/2002 HBSC survey was carried out in the following countries and regions in the WHO regions of the Americas and Europe: Austria, Belgium (the Flemish- and French-speaking populations), Canada, Croatia, the Czech Republic, Denmark, England, Estonia, Finland, France, Germany, Greece, Greenland, Hungary, Ireland, Israel, Italy, Latvia, Lithuania, Malta, the Netherlands, Norway, Poland, Portugal, the Russian Federation, Scotland, Slovenia, Spain, Sweden, Switzerland, The former Yugoslav Republic of Macedonia, Ukraine, the United States of America and Wales.

A regional sample was selected in Germany (Berlin, Hessen, North Rhine-Westphalia and Saxony). Separate studies cover the Flemish- and French-speaking populations in Belgium, and England, Scotland and Wales in the United Kingdom. As the population of Greenland is relatively small, a census of the school population was taken and all children registered at school within the target age groups were surveyed, except those not present on the day of data collection. Fieldwork took place in all countries between autumn 2001 and spring 2002. Unfortunately, the sample size for Slovakia was considered too small to be included in the 2001/2002 international data file.

About 1500 respondents in each of the three age groups were targeted in every country. Children were selected using a clustered sampling design, where the initial sampling unit was either the school class or the school. The latter was sampled when class lists were not available. The requirement for minimum recommended sample size was met in the majority of countries and regions.

While sample design is important, many other factors should be considered when looking at the findings presented here. The survey encompassed many different schooling systems and many cultures and languages; the methods of survey administration also varied within and between countries and regions. The importance of culture and language should not be underestimated when definition of concept is required as part of any item used in the survey: for example, when providing common definitions of physical activity or bullying. Samples may also differ in terms of variables such as age, socioeconomic status, school system and geographical coverage. School attendance may vary, which has the potential to introduce bias into the data presented. For example, past research has indicated that absentees are more likely to smoke and that the number of smokers may therefore be underrepresented in the data *(6)*. Further, seasonal differences in the timing of fieldwork may affect particular variables, such as diet and physical activity. Samples may therefore vary in both geographical and temporal terms, and this should not be overlooked *(7)*.

Source:
http://www.hbsc.org/

Example: National Crime Victimization Survey

Table 1.1. Example Survey: National Crime Victimization Survey (NCVS)

Sponsor	US Bureau of Justice Statistics
Collector	US Census Bureau
Purpose	Main objectives are to: • develop detailed information about the victims and consequences of crime • estimate the number and types of crimes not reported to the police • provide uniform measures of selected types of crimes • permit comparisons over time and by types of areas
Year Started	1973 (previously called the National Crime Survey, 1973–1992)
Target Population	Adults and children 12 or older, civilian and noninstitutionalized
Sampling Frame	US households, enumerated through counties, blocks, listed addresses, lists of members of the household
Sample Design	Multistage, stratified, clustered area probability sample, with sample units rotating in and out of the sample over three years
Sample Size	About 42,000 households – 76,000 persons
Use of Interviewer	Interviewer administered
Mode of Administration	Face-to-face and telephone interviews
Computer Assistance	Paper questionnaire for 70% of the interviews, both face-to-face and telephone interviews; computer assistance for 30% of the interviews
Reporting Unit	Each person age 12 or older in household reports for self
Time Dimension	Ongoing rotating panel survey of addresses
Frequency	Monthly data collection
Interviews per Round of Survey	Sampled housing units are interviewed every six months over the course of three years
Levels of Observation	Victimization incident, person, household
Web Link	*http://www.ojp.usdoj.gov/bjs/cvict.htm*

Example: National Crime Victimization Survey

I'm going to read some examples that will give you an idea of the kinds of crimes this study covers.

As I go through them, tell me if any of these happened to you in the last 6 months, that is since _ _(MONTH) _ _(DAY), 20_ _.

Was something belonging to YOU stolen, such as:

a) Things that you carry, like luggage, a wallet, purse, briefcase, book
b) Clothing, jewelry, or calculator
c) Bicycle or sports equipment
d) Things in your home – like a TV, stereo, or tools
e) Things from a vehicle, such as a package, groceries, camera, or cassette tapes

Example: National Crime Victimization Survey

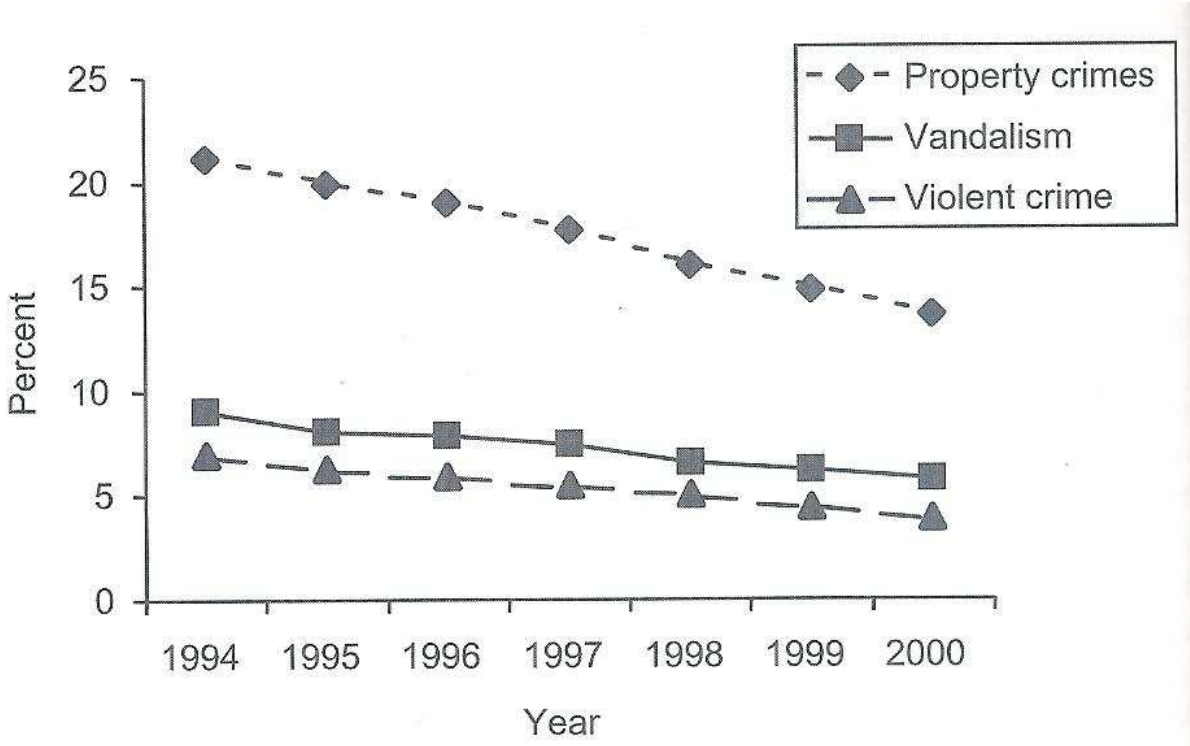


Figure 1.1 Percentage of US households experiencing a crime, by type of crime by year, based on the 1994–2000 National Crime Victimization Survey. (Source: www.ojp.usdoj.gov/bjs/)

Example: National Assessment of Educ. Progress

Table 1.4. Example Survey: National Assessment of Educational Progress (NAEP)

Sponsor	National Center for Education Statistics, US Department of Education
Collector	Westat
Purpose	Main objectives are to: • assess the academic performance of fourth, eighth, and twelfth graders in a range of subjects • reflect current educational and assessment practices • measure change over time
Year Started	1969
Target Population	National NAEP – schoolchildren in grades 4, 8, and 12 State NAEP – schoolchildren in grades 4 and 8
Sampling Frame	US elementary and secondary school children, through US counties or groups of counties, listed schools and students within schools
Sample Design	Multistage, stratified clustered area probability sample of primary sampling units (PSUs); sample of schools within PSU drawn, classrooms of students
Sample Size	2,000 schools and 100,000 students (National NAEP) 100 schools and 2,500 students per subject grade (State NAEP sample size per state) [typical sample sizes]
Use of Interviewer	None; self-administered background questionnaires completed by students, teachers, and principals; cognitive assessments completed by students; proctored by survey administrators
Mode of Administration	Paper-and-pencil self-administered questionnaires and cognitive assessment instruments
Computer Assistance	None
Reporting Unit	Students, teachers, school principals
Time Dimension	Repeated cross-sectional survey
Frequency	Conducted annually
Interviews per Round of Survey	One
Levels of Observation	Student, class, school
Web Link	*http://nces.ed.gov/nationsreportcard/sitemap.asp*

Example: National Asssessment of Educational Progress

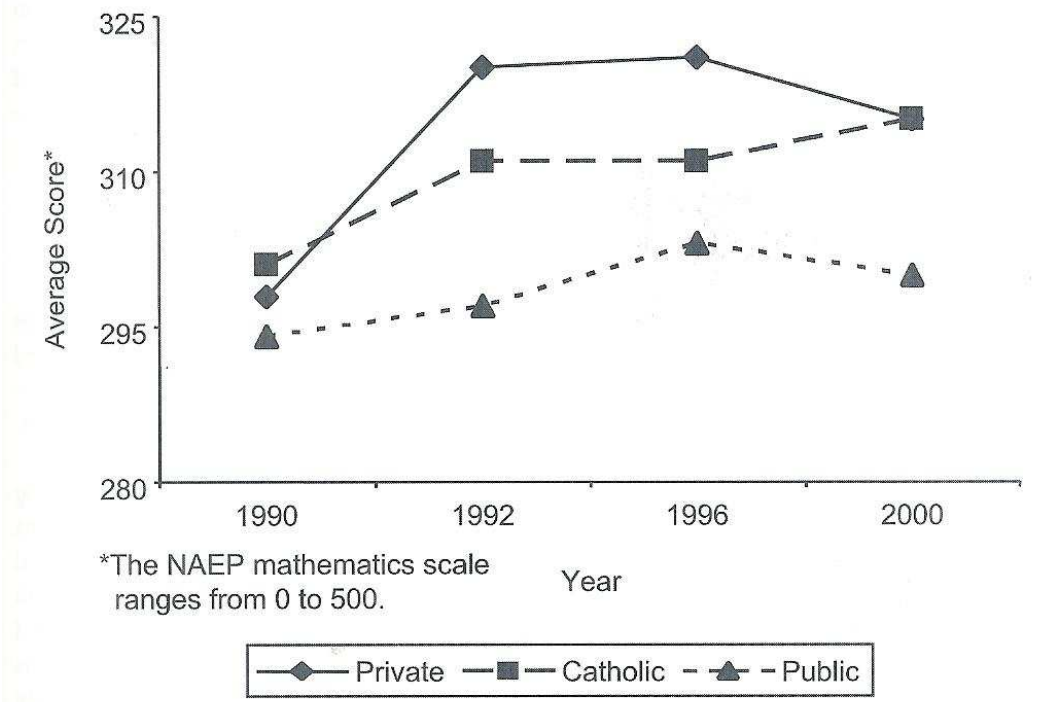


Figure 1.4 Average scale scores on grade 12 mathematics assessment, by year by type of school. (Source: US Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992, 1996, and 2000 Mathematics Assessments.)

For Next Time…

- Read:
  - Groves et al: Chapter 1
  - Groves et al: Chapter 2

- Study for "Review Quiz": Topics from 201/202 or the equivalent intro stat course that you took
  - This is diagnostic – are you ready for this course?