11. Understanding randomness

- It's not easy being random.
  - Physical devices.
  - Random phenomena in nature; e.g., radioactive decay.
  - Pseudorandom numbers on computers.

- Practical randomness: modeling variability, decision making.

- Simulation: component, trial, response variable.
  - Coin tossing: How many heads in 100 tosses of a fair coin.
  - Basketball: How many free throws before first miss, given 80% success rate. How many baskets in five throws?
  - Randomly sample three students from class of 15 women and 8 men. How likely are we to get all women?

Well, by this time I was fed up
At tryin' to make a stab
At bringin' back any help
For my friends and Captain Arab
I decided to flip a coin
Like either heads or tails
Would let me know if I should go
Back to the ship or back to jail
So I hocked my sailor suit
And I got a coin to flip
It came up tails
It rhymed with sails
So I made it back to the ship.

from "Bob Dylan's 115th Dream", 1965

12. Sample surveys

Edmd: "*You* are speaking of London. *I* am speaking of the nation at large."
Mary: "The metropolis, I imagine, is a pretty fair sample of the rest."
From *Mansfield Park*, Jane Austen, 1814.

Main ideas:

1. Examine a part of the whole: a sample is a (representative) subset of the population.

2. Randomize: Stir the pot. Why not match sample to population?

3. Its the sample size that's important, not the fraction of the population sampled.

- Does a census make sense? Cost, complexity, time, undercount.

- Population parameters and sample statistics.
  - A parameter is used in a model for a population.
  - A statistic is any summary found from data in a sample.
  - Parameters are unknown and often unknowable; e.g., μ, σ, β, ρ.
  - Statistics are known once the data are observed; e.g., $\bar{y}, s, b, r$.

- Simple random sample (SRS) of size n: each subset of n elements in the population has the same chance of being selected.
  - Need a sampling frame (list of all elements in the population).
  - Sampling variability, but samples tend to be fairly representative.

- Stratified sampling: divide population into homogeneous groups (strata) and obtain an SRS from each group.

- Cluster and multistage sampling. Divide population into representative clusters, select one or a few clusters, SRS in each.

- Systematic samples: e.g., every 10th person in list (alphabetical order).

- For a valid survey:
  - Know what you want to know.
  - Use the right frame.
  - Tune your instrument; e.g., keep questionnaire short.
  - Ask specific rather than general questions.
  - Ask for quantitative results when possible.
  - Be careful in phrasing questions: subtle differences can affect response.
  - Be careful in phrasing answers.
  - Carry out a pilot study.

- Bias is any systematic failure of a sampling method to represent its population. It's almost impossible to recover from bias. Sources of bias:
  - Convenience sample (straw poll). Captain Louis Renault: "We need a poll to estimate support for Vichy. Round up the usual suspects."
  - Bad sampling frame and other sources of undercoverage.
  - Voluntary response bias: individuals decide on their own whether to participate. E.g., internet news sites, National Household Survey.
  - Nonresponse bias; e.g., not at home, filter calls with Caller ID.
  - Response bias: anything in the survey design that influences responses, such as wording or location of questions, tone of interviewer.

- Distinguish between bias (bad) and sampling variability (inevitable but manageable).

Changes to 2011 Census announced by Canadian Government, June 2010.

- In the past the census included a short-form questionnaire sent to all households and a long-form questionnaire sent to 20% of households. Both were mandatory, with a potential fine or jail sentence for noncompliance (but jail was never imposed).

- Starting in 2011, information previously collected by the long-form census questionnaire is now collected as part of the new voluntary National Household Survey sent to 33% of households.

- Reasons given for change: privacy issues, penalties too harsh, cost.

- Industry Minister Tony Clement gave the impression that the change was supported by Statistics Canada. That was later contradicted by former Chief Statistician Munir Sheikh following his resignation July 21, 2010.

- The response rate for the mandatory 2006 long-form census was 93.7%. Statistics Canada estimated a response rate of about 50% for the voluntary survey in 2011, or 65% to 70% if additional funding were provided to encourage compliance.

- The Privacy Commissioner of Canada, who was not consulted about the change, indicated that only three complaints were laid about any aspect of the census in the last decade.

- Micheal Vonn (BC Civil Liberties Association) said that the census is not even on the list of the serious and urgent privacy issues. Airport scanners, centralized health records, and the government's Financial Transactions and Accounts Analysis Centre are much greater privacy concerns.

- James Turk (CAUT) said that the decision will result in a dramatic decline in the quality of economic and social data.

- Michael Ornstein (CAUT) said that the burden of filling out the 40-page questionnaire has been colossally inflated and the government's argument has focused on costs, ignoring benefits.

- As of Aug 28, 2010, more than 300 organizations had publicly denounced the changes, saying a voluntary survey will produce skewed results and undermine the statistical backbone of programs, businesses, and municipalities across Canada.

- Questions:
  - Main type of bias? Does the increase from 20% to 33% help?
  - Should politicians overrule experts on issues of scientific methodology? See: Indiana Pi Bill.
  - As Canadian citizens, what do we owe our country?
  - Was there a hidden agenda?

Smokers regain appeal despite anti-tobacco ads. Edm Jrnl 18 Aug 2008.

- In 2005, Ipsos-Reid asked 884 Canadian adults for Lavalife if they would date a smoker. Fully 56 per cent said they would not.

- This year, the dating website asked the same question of 6,313 Canadian adults and found the proportion of people unwilling to start a relationship with a smoker had dropped nine points to 47 per cent – 43 per cent of men and 51 per cent of women.

- The latest Canadian Tobacco Use Monitoring Survey shows smoking rates in Canada have declined from 35% in 1985 to 19% in 2006.

- The 2005 survey is considered to be accurate with 3.3 percentage points, 19 times out of 20. The 2008 survey is considered accurate within 1.23 percentage points, 19 times out of 20.

13. Experiments and observational studies

- Observational studies.
  - Retrospective studies.
  - Prospective studies.
  - Limitations, lurking variables.

- Randomized comparative experiments.
  - Four principles: control, randomization, replication, blocking.
  - Gold standard for demonstrating cause-and-effect relationships.
  - Statistical significance.
  - Compare treatment with control, blinding, placebo.
  - Confounding.

- Observational study:
  - May involve a random sample, but usually not.
  - Look for associations between explanatory variables and response (outcome) variables, often over a period of time.
  - Difficult to assert causal effect due to lurking variables.
  - Matching sometimes employed to reduce variability.

- Retrospective study: Identify subjects in groups determined by responses of interest (e.g., cancer patients, healthy individuals). Then look for possible explanatory variables (life style, exposure, location, work, family). Useful when one outcome is rare. Wrong conditional distributions.

- Prospective study: Identify subjects and follow over a period of time. Observe explanatory and response variables. More costly than a retrospective study, but better data.

Running can slow the effects of aging. Edmonton Journal, Aug 14, 2008.

- Starting in 1983, Stanford University, participants over age 50
- 538 members of a running club, 423 healthy non-runners
- Yearly questionnaires about ability to perform everyday activities
- Over time, both groups became more disabled
- On average, onset of disability started 16 years later for runners
- Gap between groups increased over time
- Runners had less heart disease
- Similar rates of osteoarthritis and knee replacements (unexpected)

- After 19 years, 15% of the runners had died, 34% of the nonrunners.
- Researchers project four-year difference in survival once all have died.
- In 1983, running times averaged 4 hours per week.
- By 2004, times averaged 76 min per week, health benefits still observed.
- The researchers controlled for differences between the two groups in age, sex, smoking history, body mass index, initial disability, and weekly aerobic exercise.
- Questions:
  - Main conclusion valid?
  - Explanatory and response variables?
  - What about drop-outs?
  - How does one "control for differences"?

Do we really know what makes us healthy?

You may be interested in this article by Gary Taubes published in the New York Times, Sept 16, 2007. The article discusses why causal effects suggested in observational studies are often not confirmed or are even contradicted in subsequent clinical trials. A main focus of the article is Hormone Replacement Therapy (HRT). Reports in 1985 from the Nurse's Health Study (a large prospective observational study) suggested that HRT lowers the risk of heart attacks. A subsequent clinical trial found that HRT actually increases the risk of heart attacks among older women. Taubes discusses the difficulties in drawing valid conclusions about cause-and-effect relationships from observational studies and recommends skepticism when reading reports from such studies. To see the article, follow the link at the site below:

`http://www.stat.ualberta.ca/~hooper/teaching/stat141/GaryTaubes.htm`

Randomized, comparative experiments

- Identify a set of experimental units (subjects, participants, lab rats).
  - The set of units is usually not a random sample.
- Identify one or more explanatory variables, called factors.
  - The specific values of factors are called levels.
  - A treatment is a combination of factor levels.
- Identify one or more response variables.
- Allocate units to treatments, then (eventually) observe responses.
- An important goal is to determine whether differences observed among responses for different treatments are *statistically significant*; i.e., whether the differences might have occurred by chance.

Four principles.

1. Control sources of variation other than the treatment factors.

2. Randomly (not haphazardly) allocate units to treatment groups.
   This is *not* the same as randomly sampling units from a population.

3. Two kinds of replication.

   - Each level of each factor is applied to multiple units. This allows an evaluation of variability and hence of statistical significance.
   - The entire experiment is replicated at another time and/or location.

4. If appropriate, arrange the units into more homogeneous *blocks* and randomize the allocation of units separately within each block. This allows more accurate assessment of treatment differences by making comparisons within blocks, where the variation is less.

Example: Tomato experiment (in text).

- Purpose: investigate effect of fertilizer on taste of tomatoes.
  - Unit = tomato plant, $n = 18$.
  - Factor = amount of fertilizer (low, medium high).
  - Response = average taste rating (1 to 7) by panel of judges.

- Completely randomized design.

- Suppose 6 tomatoes were purchased from each of 3 stores.
  Randomized block design.

- Add a second factor: amount of water (low, high).
  Completely randomized design.
  Randomized block design.

- When the levels of two factors are associated in such a way that their effects cannot be separated, we say the two factors are *confounded*.

- Given appropriate randomization, treatment allocation is independent of any lurking variables. However, confounding can still occur due to (possibly unconscious) attitudes about the treatments.

- An individual associated with the experiment who is not aware of the treatment assignments is said to be blinded.

- There are two classes of individuals who can affect the outcome of an experiment:
  - Those who could influence how the subject responds to the treatment (subjects, treatment administrators, technicians).
  - Those who evaluate the responses (judges, treating physicians).

- Single and double blinded experiments.

- In clinical trials, a new treatment is usually compared with the best treatment currently available. Subjects assigned to the latter are said to be in the control group. Ethical issues, equipoise.

- If the currently best treatment is to do nothing, then subjects in the control group are usually assigned a placebo; i.e., something resembling the new treatment but that has no effect (beyond a psychological effect). This allows subjects to be blinded.

- Confounding often occurs when the experiment is not properly randomized. E.g., clinical trials for the Salk polio vaccine. Two designs were employed. In the first, parents were allowed to opt out and have their children placed in the control group. In the second, school districts were randomly allocated to treatment and control groups.

Diners spend more if menu avoids $ sign. Edmonton Journal, Aug 14, 2008.

- A team at Cornell University tested whether the price presentation on menus affected how much customers were willing to spend.

- The researchers presented three versions of a typical lunch menu at a local "upscale-casual" restaurant, and gleaned their results from 201 participants who used a menu with prices listed as XX, $XX or scripted words; e.g., 20, $20 or twenty dollars.

- They found that the XX menus yielded, on average, $5.55 more in spending than the other two formats (which had about equal results).

- To appear in the *International Journal of Hospitality Management*.

- How would you design this experiment?

- Does wearing high heels increase the risk of knee problems (osteoarthritis) in later life? A study was reported on the CBC radio program "As it happens" on January 7, 2004. The study involved two groups of women, 29 awaiting knee surgery and 85 with no knee problems, all in their 50's or 60's. The investigators determined, for each group, the proportion of women who had previously worn high heels. Contrary to expectation, they found a slightly lower proportion in the group awaiting knee surgery.

  a) Is this an experiment or an observational study? Describe the design.
  b) What do you conclude from the study?
  c) How were the groups selected? Was randomization used?
  d) Sources of bias?

Finger length can predict math and reading skills

- From the Ottawa Citizen, June 2007, a report about an article in the British Journal of Psychology by Brosman et al. at University of Bath.

- Higher levels of prenatal exposure to testosterone are known to produce longer ring finger relative to index finger.

- Testosterone is also thought to affect brain development. Higher levels encourage growth in right half of brain, associated with spatial and math skills, while slowing development of left half, associated with language.

- Study examined 75 six & seven year olds (33 boys and 42 girls) in southwest of England. From photocopies of hands, measured "digit ratio" = DR = (length of index finger)/(length of ring finger). Measure from crease to tip.

- DR remains $\approx$ fixed after age five. Average DR among adults is 1.00 (women) and 0.98 (men). Within study: 0.96 (girls) and 0.95 (boys).

- Study compared DR with scores on numeracy and literacy tests. Children with DR < 1 were better at math than language. And vice versa.

- For boys, study found link between low DR and higher math scores. Girls with higher DR had better literary scores.

- Caveat: DR is treated as a surrogate for level of prenatal exposure to testosterone. DR is not a measure of intelligence. DR may indicate predisposition for math or language.

- Why not use difference instead of ratio?

Women found less likely to sacrifice for love

- From Edmonton Journal, 30 Aug 2007, regarding paper by Catherine Mosher & Sharon Danoff-Burg in *Gender Issues*.

- Men are more likely than women to place romance above career, education, and travel.

- Questionnaires were given to 237 college students, 80 men and 157 women. Typical question: "Suppose that a close romantic relationship developed ... If you had to choose between the relationship and having a professional career, which would you choose?"

- A "relational score" was calculated for each subject (apparently with lower scores indicating greater willingness to sacrifice for love). Means and standard deviations: for men (1.39, 0.27), for women (1.46, 0.24).

- The authors were surprised by result. Mosher noted: "What they state on paper may not necessarily translate into what they do down the road. It would take longitudinal research to find out. I would have to look more at the sociological literature, at any changes and the degree to which men are involved in their relationships versus their work, to start to look at generational differences."

- A referee was not surprised: "We know that men fall in love faster than women do, and that sort of thing."

- Bias? How was the "score" formulated?

- Statistical significance? Later.

Wonder drug helps people lose weight, stop smoking

- From an article in Edmonton Journal, August 30, 2004.
- A once-a-day pill that helps people to lose weight and stop smoking could be available within two years.
- The drug rimonabant works on a newly discovered system in the brain affecting control of appetite and urge to smoke.
- Being overweight, smoking, and diabetes are all factors in heart disease.
- The drug works on the endocannabinoid system in the brain, which researchers believe plays a vital role in how the body regulates weight.

- Latest results from a year-long study, issued at the European Society of Cardiology meeting in Munich, show that 40 to 45 per cent of overweight and obese people in the trial lost 10% of their weight.

- Results from the European study of 1500 people found that those on a 20 mg dose lost 6.6 kg in a year, those on 5 mg lost 3.4 kg, and those on a placebo lost 1.8 kg.

- All those in the trial followed a diet reducing their calories by 600 a day and were advised to take exercise.

- Across the world, 13000 people are involved in seven rimonabant trials which are looking at its effect on weight, smoking, and diabetes.

- Update from Wikipedia: Approved by EU in 2006. Approval for USA denied by FDA in 2007 due to concerns about safety.