Lecture 1

Brian Caffo

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Johns Hopkins University

September 5, 2007
Table of contents

1. Table of contents
2. Outline
3. Biostatistics
4. Set notations and experiments
5. Probability
Outline

- Discuss biostatistics
- Cover syllabus
  - mathematical prerequisites
  - web site
  - quiz and homework schedule
  - test schedule
  - R
- Abstract the idea of an experiment
- Develop basic set theory to be used in the development of probability
- Start discussing probability
Biostatistics defined

From the Johns Hopkins Department of Biostatistics 2007 self study:

*Biostatistics is a theory and methodology for the acquisition and use of quantitative evidence in biomedical research. Biostatisticians develop innovative designs and analytic methods targeted at increasing available information, improving the relevance and validity of statistical analyses, making best use of available information and communicating relevant uncertainties.*
Example: mortality count

• In October of 2004 and again in 2006 researchers Roberts and Burnham from the Johns Hopkins Bloomberg School of Public Health published landmark studies estimating the number of excess deaths in Iraq after US military operations.

• The second published number of 654,965 excess deaths prompted the following:

No, I don’t consider it a credible report. Neither does General Casey and neither do Iraqi officials. I do know that a lot of innocent people have died, and that troubles me and it grieves me. And I applaud the Iraqis for their courage in the face of violence. ... No question, it’s violent, but this report is one – they put it out before, it was pretty well – the methodology was pretty well discredited. - GW Bush October 11th.
The Canadian National Breast Cancer Screening study found no benefit of early tumor detection via digital mammography for women aged 40-49 in a large randomized screening trial, contradicting standard practice of radiology at the time. Gray, in a 2003 Canadian Medical Association commentary, states that criticisms of the study focus on design, methodology and conclusions.
Example: hormone replacement therapy

A large clinical trial (the Womens Health Initiative) published results in 2002 that contradicted prior evidence on the efficacy of hormone replacement therapy for post menopausal women and suggested a negative impact of HRT for several key health outcomes. Based on a statistically based protocol, the study was stopped early due an excess number of negative events.
Example: ECMO

In 1985 a group at a major neonatal intensive care center published the results of a trial comparing a standard treatment and a promising new extracorporeal membrane oxygenation treatment (ECMO) for newborn infants with severe respiratory failure. Ethical considerations lead to a statistical randomization scheme whereby one infant received the control therapy, thereby opening the study to sample-size based criticisms.
Summary

• These examples illustrate the central role that biostatistics plays in public health and the importance of performing design, analysis, and interpretation of statistical data correctly.

• At the Johns Hopkins SPH, a consistent philosophy for conducting biostatistics includes:
  • A tight coupling of the statistical methods with the ethical and scientific goals.
  • Emphasizing scientific interpretation of statistical evidence to impact policy.
  • Acknowledging and assumptions and evaluating the robustness of conclusions to them.
Experiments

Consider the outcome of an experiment such as:

- a collection of measurements from a sampled population
- measurements from a laboratory experiment
- the result of a clinical trial
- the result from a simulated (computer) experiment
- values from hospital records sampled retrospectively
- ...
The sample space, $\Omega$, is the collection of possible outcomes of an experiment.

Example: die roll $\Omega = \{1, 2, 3, 4, 5, 6\}$

An event, say $E$, is a subset of $\Omega$.

Example: die roll is even $E = \{2, 4, 6\}$

An elementary or simple event is a particular result of an experiment.

Example: die roll is a four, $\omega = 4$

$\emptyset$ is called the null event or the empty set.
Interpretation of set operations

Normal set operations have particular interpretations in this setting

1. $\omega \in E$ implies that $E$ occurs when $\omega$ occurs
2. $\omega \notin E$ implies that $E$ does not occur when $\omega$ occurs
3. $E \subset F$ implies that the occurrence of $E$ implies the occurrence of $F$
4. $E \cap F$ implies the event that both $E$ and $F$ occur
5. $E \cup F$ implies the event that at least one of $E$ or $F$ occur
6. $E \cap F = \emptyset$ means that $E$ and $F$ are mutually exclusive, or cannot both occur
7. $E^c$ or $\bar{E}$ is the event that $E$ does not occur
Fun aside: Russell’s paradox

Russell’s paradox is one of the most famous results of set theory

- Consider, $R$, the set containing all sets that do not contain themselves as an element
  
  Alternatively, consider writing down a catalog of all catalogs who do not have themselves listed as an entry

- Does $R$ contain itself?
  
  → If yes, then $R$ is not allowed to be in $R$, by the definition
  
  → If no, then $R$ has to be in $R$, by the definition
Set theory facts

- DeMorgan’s laws
  \[(A \cap B)^c = A^c \cup B^c\]
  \[(A \cup B)^c = A^c \cap B^c\]

Example: If an alligator or a turtle you are not \([(A \cup B)^c]\) then you are not an alligator and you are also not a turtle \((A^c \cap B^c)\)

Example: If your car is not both hybrid and diesel \([(A \cap B)^c]\) then your car is either not hybrid or not diesel \((A^c \cup B^c)\)

- \((A^c)^c = A\)

- \((A \cup B) \cap C = (A \cap C) \cup (B \cap C)\)
Probability: some discussion

- Useful strategy used in much of science: For a given experiment
  - attribute all that is known or theorized to a mechanistic model (mathematical function)
  - attribute everything else to randomness, even if the process under study is known not to be “random” in any sense of the word
  - Use probability to quantify the uncertainty in your conclusions
  - Evaluate the sensitivity of your conclusions to the assumptions of your model
Probability: some discussion

- Probability has been found extraordinarily useful, even if true *randomness* is an elusive, undefined, quantity
- *frequentist* interpretation of probability
  - A probability is the long proportion of times an event will occur in repeated identical repetitions of an experiment
- Other definitions of probability exists
- There is not agreement, at all, in how probabilities should be interpreted
- There is (nearly) complete agreement on the mathematical rules probability must follow
Probability: some discussion

- An alternative interpretation of probability is so-called “Bayesian”
- Named after the 18\textsuperscript{th} century Presbyterian Minister / mathematician Thomas Bayes
- Bayesian interprets probability as a subjective degree of belief
  - For the same event, two separate people could have differing probabilities
  - Bayesian interpretations of probabilities avoid some of the philosophical difficulties of frequency interpretations