Topic 4
Producing Data
Preliminary Steps

Outline

Getting Started

Begin with a review of previous work and exploratory data analysis, looking for

- patterns and associations,
- inconsistencies in the scientific literature, and
- examine the data using quantitative methods.
  - summary statistics for quantitative variables, tables for categorical variables
  - graphical methods - boxplots, histograms, scatterplots, time plots for quantitative data - bar charts for categorical data

Getting Started

The strategy of these investigations is frequently the same - look at a sample in order to learn something about a population or to take a census of the total population. Designs for producing data begin with some basic questions:

- What can I measure?
- What shall I measure?
- How shall I measure it?
- How frequently shall I measure it?
- What obstacles do I face in obtaining a reliable measure?

Getting Started

The frequent goal of a statistical study is to investigate the nature of causality

- Try to explain the values of some response variables based on knowing the values of one or more explanatory variables.
- The associated phenomena could be caused by a third, previously unconsidered factor, called a lurking variable or confounding variable.

  To mitigate the impact of confounding.
  - Subdividing the population under study into smaller more similar groups that are more similar. This subdivision is called cross tabulation or stratification.
  - Developing a mathematical or probabilistic model. These models often take the form of a mechanistic model that takes into an account the variables in the cross tabulation and builds a parametric model.

Ethical Issues

As a citizen, we should participate in public discourse. Those with particular training have a special obligation to bring to the public their special knowledge. Such public statements can take several forms.

- We can speak out as a member of society with no particular basis in our area of expertise.
- We can speak out based on the wisdom that comes with this specialized knowledge.
- We can speak out based on a formal procedure of gathering information and reporting carefully the results of our analysis.

We are obliged to be clear about the nature of our communication, following the highest ethical standards. In the same vein, as a consumer of information, we should have a clear understanding of the perspective in any document that presents statistical information.

Ethical Issues

Professional statistical societies have documents that provide guidance on what can be sometimes be difficult judgements and decisions.

Two sources of guidance are

- the *Ethical Guidelines for Statistical Practice* from the American Statistical Society and

  `http://www.amstat.org/about/ethicalguidelines.cfm`

- the International Statistical Institute *Declaration on Professional Ethics*

  `http://www.isi-web.org/about-isi/professional-ethics`

Ethical Issues

Exercise. Read the American Statistical Association *Ethical Guidelines for Statistical Practice*

`http://www.amstat.org/about/ethicalguidelines.cfm`

and give brief answers to the following questions.

1. Give five issues that you should examine in a research paper to see that it has complied with the *Ethical Guidelines*.
2. Give five issues that you should consider before joining a research project.
3. Give five ethical obligations that a person has even though they are not the statistician on the project.
4. If you witness a breach of ethical standards, what are some that the actions you should consider.