STAT 500

Observational Studies

Observational Studies

- In some cases, the treatments cannot be assigned to experimental units by some rule.
  - For example, study of the effects of smoking on cancer with human objects as the experimental units
  - Neither ethical nor possible
- We can still gather data by observing members of the target population as they naturally exist.
- This type of study is called observational study and is not an experiment.
- We can still analyze the data from observation studies, but cannot draw conclusion of causation.

Observational Studies

- Gather information through
  - Census: Observe all members of population
  - Haphazard (convenience) sample
  - Representative random sample
- Inferences
  - Only about associations, not prove causation
  - Only using representative random samples

Simple Random Samples

- Simple random sample of size n without replacement:

  every subset of n unique units has the same probability of being selected - more typical

- Simple random samples of size n with replacement:

  on each draw every member of the population has the same chance of being selected and the selected unit is put back into the population before the next unit is selected (some units may be selected more than once)

Simple Random Samples

- Large Populations - Small Samples
  - Simple random sample without replacement is similar to simple random sample with replacement.
  - Selection of sample unit changes probability of selection of another sample unit by a very small amount.
  - Treat two sampling schemes in the same way.

Random Samples

- Only consider simple random samples, but there are many other sampling schemes that produce representative samples (Stat 521: Survey Sampling)
- The sampling procedure dictates the method of analysis
- Can make predictions and inferences about associations
- Causal inferences are not justified

Problems outside Sampling

- Non-response bias
  - Sampling units do not respond to survey
  - Low response rate leads to convenience sample
  - Methods for increasing response rate - initial introduction, incentives, multiple reminders
- Response bias
  - Non-truthful responses to survey questions
  - Faulty memory, lack of understanding of questions, omissions

Problems outside Sampling

- Wording of Questions
  - Poor wording - confusing
  - Leading questions

Components of Observational Studies

- Clear statement of research question(s) and objective(s)

- Identification of target population(s)

- Identification of sampling units
  - Members of the population who provide the measured response

- Measurable characteristics of sampling units (factors)
  - Features of the sampling units
  - Analyze associations with a specified outcome

- Specification of the sampling procedure dictates methods of analysis and restricts types of inference

Observational Studies - Examples

- Retrospective study of potential effects of smoking on lung cancer
  - Simple random sample of patients diagnosed with lung cancer at some specific set of hospitals
  - Independent simple random sample of non-lung cancer patients from the same hospitals
  - Compare smoking histories for the two samples

- Prospective study: Nesting success of pheasants
  - Random sample of N locations
  - Find nests and implant transmitters in chicks
  - Relate survival probability to features of the surrounding habitat

Observational Studies - Examples

- Prospective study: Nurses Health Study
  - About 10,000 nurses volunteered to enroll (females who were 20-30 years old at the start)
  - Food intake diaries
  - Examine association between fat intake and incidence of heart disease
  - No control of other factors that might affect incidence of heart disease (genetics, exercise, weight, stress ...)
  - Useful information on associations, but cause and effect inferences can not be justified

Another Example: Fluoridation of Water Supplies and Cancer

- Data collection
  - 10 largest US cities with water fluoridated starting 1951-1956
  - 10 largest US cities that were not fluoridated by 1969
- Cancer deaths per 100,000 population

	1950	1970	change
Fluoridated	180	217	+37
Non-Fluoridated	178	197	+19

- Effect of fluoridation? Only if assume no other difference between cities.

Another Example: Fluoridation of Water Supplies and Cancer

Fluoridation of Water Supplies and Cancer-A Possible Association? by Oldham and Newell (1997), Applied Statistics

- Oldham and Newell's analysis showed that the two groups of 10 cities differed in their age-sex-race structure in 1950.

- By 1970, the two sets of cities differed much more in their demographic structure than they had in 1950.

- When these demographic changes are taken into account, Oldham and Newell's analysis showed that "excess" cancers increased 1% in fluoridated cities, 4% in non-fluoridated cities.

Analysis of Observational Studies

- Data analysis may be based on the sampling distributions of summary statistics, such as sample means

- Data analysis may be based on assumptions about population distributions
  - We will mostly consider analyses for data sampled from populations with normal distributions
  - There are many other distributions, such as the Weibull distribution for survival times

- Central Limit Theorem: The sum (or average) of a large number of independent observations is approximately distributed as a normal random variable