Lecture 3
Outline

Summary of Previous Lecture
  Normal Probabilities

Normal Distribution
  Central Limit Theorem and Applications

Sampling and Study Design
  Population, Sample and Inference
  Population
  Sample
Part I

Review of Previous Lecture
Review of Previous Lecture

- Know what you are talking about – define the population.
- Make others trust what you are talking about – avoid bias in sampling.
Part II

Population, Sampling and Study Designs
Generating Random Real Numbers and Integers

- Values from any random distribution can be constructed from random Uniform(0,1) variates
- Excel: `randbetween (2, 10), randuniform (0.2, 0.8)`.
- SAS: `uniform (0)` gives a number in (0, 1].

```plaintext
data ran;
    x = uniform (0);
    y = int (uniform (0) * 10 + 1);
    if uniform (0) > 0.5 then z = 1;
    else z = 0;
proc print data = ran;
run;
```
Sampling Techniques

Definitions

- A *simple random sample* (SRS) of size $n$ from the population is chosen in such a way that every set of $n$ individuals has an equal chance of being selected.

- To select a *stratified random sample*, first divide the population into groups of similar individuals, called *strata*. Then choose a separate SRS from each stratum and combine them.
The Right and The Answerable Question

- Experimental units or subjects.
- The response (outcome, endpoint) variable: actual or "surrogate"?
- The explanatory (treatment, risk factor) variable(s).
- Estimation or Testing? (Null hypothesis)
- How accurate is accurate? How different is different? (Alternative hypothesis)
Observation and Experimentation

Definitions

- An *observational study* observes individuals and measures variables of interest, but does not attempt to influence the responses (anecdotal, ecological).

- An *experiment* deliberately imposes some treatment on individuals in order to observe their response (laboratory, clinical trials).
Gentlemen, I have to report that this confounded experiment has a previously unnoticed confounding factor.
Confounding

Definitions

- A confounding factor (confounder) is associated with both the exposure and the outcome.
- When a confounder is not measured or not included in the analysis, the estimated relationship between the exposure and the outcome can be distorted...
Example

Diet and colon cancer: a confounding factor is *age*

- People in different age groups have different dietary and eating habits.
- The risk of any cancer tends to increase with age.
Observe or Experiment

- Experiments (clinical trials) are more expensive, but make it easier to control for confounding (say, through randomization).
- It is more reliable to draw inferences regarding causal relationships from experiments.
- Observational studies can be larger and more comprehensive.
- Often it is not possible to do an experiment due to ethical or practical reasons.
Cohort

Definition
A *cohort* of people is a group of people for which membership in the group is clearly defined.

Examples

- All students enrolled in this course.
- Men and women between the ages of 30 and 62 from the town of Framingham, MA in 1948. (Framingham Heart Study)
Definition

- A cohort with known levels of exposure to potential causative and protective factors is recruited and followed over time to determine who experiences the outcome of interest (typically disease onset or death).

- A cohort study can be *retrospective* if the cohort and the outcome are identified/ascertained from historical records.

- Sometimes members in the cohort that have different levels of exposure are compared with each other. Alternatively, the outcome rate in the cohort can be compared to standard population rates.
Characteristics of a Cohort Study

- Typically long-term studies.
- More expensive than case-control studies, but can be used to study more than one disease outcome.
- Biases are less of a problem in cohort studies (in a sense, a cohort can be considered to be a population.)
- Not suitable for rare outcomes.
- More difficult to study the effect of unknown risk factors.
Case-Control Study

Definition

► In a *case-control study*, subjects are ascertained based on whether they have experienced the *outcome* of interest (the *cases*) or not (the *controls*). The prior histories of the cases and controls are then compared.

► Case-control studies are usually retrospective.
Characteristics of Case-Control Study

- Relatively short term studies.
- Less expensive than cohort studies, but can only be used to study one outcome at a time.
- Preferred design when the outcome is rare (ethics issues here with a clinical trial)
- Can be used to study unknown risk factors.
- More susceptible to bias: differential survival, differential memory of exposure.
**Definition**

A comparative study compares two or more techniques, treatments, or the presence versus absence of some risk factor(s).
Definitions

- A *placebo* treatment is designed to appear exactly like a comparison treatment, but be devoid of the active part of the treatment.

- The *placebo effect* results from the knowledge that one is being treated at all, rather than due to physical, physiological, or chemical activity of a treatment.
Blinded Studies

Definitions

- A study is *single blinded* if treated subjects are unaware of which treatment they are receiving.
- A study is *double blinded* if both the treated subjects and the individuals evaluating the outcome variables are also unaware of which treatment the subjects are receiving.
- In this case typically only the statisticians know which treatment ("A" or "B") is which!
Comparable Groups

- Ideal comparison: take both the red and the blue pill.
- Twin studies.
- Matched-pair: identical in all but the treatments/exposure. e.g., matched case-control study.
- Cross-over study: the same subjects take both treatments in a sequential manner, often with a “wash-out” period in between.
- Randomization: On average, the treatment groups will be similar, even with respect to variables not measured or thought about beforehand.
Helena Study

- Exposure of interest: second hand smoking.
- Real explanatory variable: smoke-free ordinance (starting 6/5/02; suspended by a legal challenge after six months)
- No individual-level exposure data.
- Outcome of interest: heart diseases.
- Real outcome: acute myocardial infarction (AMI) - AMI admissions for month.
- Historical controls.
Helena Study

- Observational.
- Ideal target population: everybody.
- Helena cohort.
Helena Study: Sampling Scheme

- Three local hospitals.
- Computer search all primary and secondary diagnosis of AMI.
- Chart review: secondary diagnosis and out of area zip codes.
- Cases are included if a) residents; b) spent the night before; c) ate at least one meal.