7.2 Sampling Plans and Experimental Design

Statistics 1, Fall 2008

Inferential Statistics

- Goal: Make inferences about the population based on data from a sample

- Examples
  - Estimate population mean, μ or σ, based on data from the sample

Methods of Sampling

- I'll cover five common sampling methods, more exist
  - Simple random sampling
  - Stratified random sampling
  - Cluster sampling
  - Systematic 1-in-k sampling
  - Convenience sampling

Simple Random Sampling

- Each sample of size n from a population of size N has an equal chance of being selected

- Implies that each subject has an equal chance of being included in the sample

- Example: Select a random sample of size 5 from this class.

Computer Generated SRS

- Obtain a list of the population (sampling frame)
- Use Excel
  - Generate random number for each subject using rand() function
  - copy|paste special|values to fix random numbers
  - sort on random number, sample is the first n in this list
- Use R (R is a free statistical software package)
  - sample(1:N,n), N=population size, n=sample size
  - Returns n randomly selected digits between 1 and N
  - default is sampling WOR

Stratified Random Sampling

- The population is divided into subgroups, or strata

- A SRS is selected from each strata

- Ensures representation from each strata, no guarantee with SRS

Stratified Random Sampling

- Example: Stratify the class by gender, randomly selected 3 female and 3 male students

- Example: Voter poll – stratify the nation by state, randomly selected 100 voters in each state

Cluster Sampling

- The subjects in the population may be grouped, or clustered, in some natural way
  - Eggs at a supermarket are clustered into cartons
  - People in the US are clustered into households
  - Students in an elementary school are clustered into classes
- Cluster sampling involves taking a SRS of clusters, then surveying *all* items in the cluster

Cluster Sampling

Example

- You have a shipment of 100 cartons, each containing a dozen eggs
- You want to estimate the proportion of broken eggs
- SRS of 10 cartons of eggs
- Survey all 12 eggs in each of the 10 cartons and note whether it is broken or not (categorical data)

Systematic 1-in-k Sample

- Easiest way to draw a sample
- Obtain an ordered list of the population
- Select a random start point from 1,2,…,k
- Select every kth item in the list after your start point
- Caution: sample could be biased
  - Example: Number of customers at a supermarket

Systematic 1-in-k Sample

Example: 1-in-6 sample from this class

- Toss die to get start point

- Use roster to select start point and every 6th student thereafter

Convenience Sample

- A sample that is simple or easy to obtain without randomization
  - Internet polls
  - Shopping mall polls
  - A poll of your 5 closest friends
- These samples are often biased
- You can't make inferences about the population from them

SRS only in this class

All the procedures we study are for SRS only

Study Design

- Observational Study vs Designed Experiment
- Treatment: any manipulation of the subject's environment
- Examples
  - Does music reduce pain for respiratory therapy patients? Treatment = music
  - Do magnets reduce pain? Treatment = magnet
  - Does exercise reduce the risk of getting a cold? Treatment=exercise

Observational Study

- Data for each subject is simply recorded or measured with no treatment being imposed

- Examples:
  - Political polls
  - Opinions polls: prestigious occupations
  - Exercise and risk of getting a cold – if you just record the amount of exercise and the number of colds for each person on a year

Designed Experiment

- Experimental treatment or condition is imposed on experimental units, then data are collected or measured

- Characteristics of a good designed experiment
  - Randomization into Control group/treatment group
  - Placebo: a fake treatment, Placebo effect
  - Double-blinding: neither subjects nor researchers know who is in the treatment/control group

Examples of Designed Expt

- Clinical trials: give patients a new drug, then observe whether they are cured or not

- Manufacturing: apply a new tool or machine, then observe whether defects are reduced

- Exercise and colds: have subjects exercise then observe the number of colds

Cause and Effect

Can only be proved in a Designed experiment!!!