6.S085 Statistics for Research Projects **IAP 2014**

Lecture 7: January 29

Lecturer: Ramesh Sridharan and George Chen *Notes by: William Li*

7.1 Experimental Design

Simple random sampling (SRS): randomly choose subset of individuals from a population (N people) without replacement

Sample n individuals: any subset of n individuals is equally likely

Blocking/controlling for confounding variables: dealing with Simpson's paradox

Randomization: need this often to get independent samples and to meet assumptions for stat. tests

Replication: good experiments should be reproducible! Replacing experiment should yield similar reuslts

Control/baseline comparison: to measure effect of treatment, need ref to compare against.

7.2 Stratified Random Sampling

Block things out, and do simple random sampling (SRS) in each block

For a stratum: how large of a simple random sampling?

Definition 7.1 (Proportional Allocation) *Number of individuals in a stratum matches stratum's relative size in population*

Definition 7.2 (Neyman Allocation) *Combines proportional allocation with looking at variances within a strata*

Higher variances → larger SRS (higher relative size of stratum in population)

Definition 7.3 (Cluster sampling) *Previous methods require sampling from whole population or every stratum*

Idea: partition population into "natural" groups (each group well-represents the population)

Cluster Sampling:

1. Randomly sample a few groups
2. For each chosen group, obtain SRS

Example: Polling a city

1. Divide city into blocks
2. Randomly choose a few blocks
3. Within each chosen block, get SRS

7.3 Problem Set Discussion

$\log y = \beta_1 \log x + \beta_0 + \epsilon$

$y \propto \exp beta_0 x^{\beta_1}$

Higher residuals