III. Sampling, Probability part 1

D.O.H. next Fri. Apr. 28 3-4 pm (Tu. cancelled)

Review

1. Potential subjects
\[ \text{randomize} \]
2. Completely randomized design

Diff. design

"pre-post design" ex. H.W. 1 #3 - person #

\[ \begin{array}{cc}
(T) & (C) \\
1.3 & 0.6 \\
1.1 & 1.1 \\
0.6 & 0.2 \\
\end{array} \]

\[ (T-C) \text{ difference} = -0.1 \]

C.S. #10 (pg. 53) "Literary Digest Poll"

Sample un-sampled

Pop. (all subjects of interest to us)

-Subset observed
-Not observed

Parameter

\[ \bar{p} = \mu = \text{mean} \]

L.D. \( n = 2,400,000 \)

\[ 9 = 4.3\% - \text{D.O.H. thinks} \]

L.D. \( \hat{p} = 4.3\% \)

Actual \( \bar{p} = 2.8\% \)

\( \delta \text{MR} = 1.5\% \)

Even w/ a sample size of \( n = 2,400,000 \)
Basic principle: how sample & unsample should be related
Want sample & unsample to be as similar as possible
in all relevant ways except for sample-unsampled distinction. = same Pop
idea as w/ Exp. design

do choose sample from pop. at random
2 ways: (1) at random with replacement (Independent)
   identically distributed (IID) sampling
   (2) at random w/out replacement (Simple
random sampling (SRS)).

Did L.D. use IID or SRS?  No
Did their method of sampling produce a sample
like what you would get from IID or SRS?  No

What they did wrong
Bias - a systematic tendency to over or under-estimate
the truth.
IID & SRS - unbiased

Selection bias - a systematic tendency to over or
under-represent relevant subsets of pop.
L.D. poll had selection bias - oversampled
on rich people, income -> politics,
& rich people tend to vote Republican
Non-response bias - people who choose to respond to
survey may be systematically different in
relevant ways.

L.D. poll may have had huge problem w/ nonresponse
bias - their response rate was only 2.4m ill. / 10 m ill.
2/4 AMS 5

A good response rate is usually 85% or higher. Response bias - the way questions are asked can influence answer.

Howthorne effect - people tend to change their behavior when they know they're being watched.

Who's watching what kind of TV shows?

Idea: