

What to know for the test on wednesday:

Chapters 1-6

- Individuals
- Variable
- Data
- Know three types of studies:
 - Observations - collect data but have not applied a treatment
 - Sample Surveys
 - Experiments - have applied some sort of treatment
 - Know the differences and similarities
 - Know what each of them are
 - Know what a census is
- Population versus sample
 - Be able to identify the sample and the population from description
 - Know the difference between the two
- Poor Sampling Methods
 - Convenience Sample
 - Voluntary Response Sample - information is requested, people self select to participate in the survey
 - Know that these two methods result in bias
- Most Basic Good Sampling Method
 - Simple Random Sample (SRS)
 - Definition
 - How to take
- Parameter - measure that we're looking for
 - generally will not know the value of a parameter
 - attempt to estimate parameter using the statistic
- Population
- Statistic
- Sample
- Bias
 - Know how to reduce bias
 - Know definition
- Variability
 - Know how to reduce variability - bigger samples
 - Know definition (how to spread out the values of the statistic are when we take many samples)
- \hat{p}
 - how to compute
 - number of successes / total number of chances
 - quick formula for MOE at 95% confidence for \hat{p}
 - $1 / \text{square root of } n$
 - know basics for confidence interval (estimate \pm or margin of error (MOE))
 - to get a smaller (narrower interval) \rightarrow larger sample size
 - to get a smaller (narrower interval) \rightarrow lower level of confidence
- Sampling errors - caused by the act of taking a sample
 - random sampling error (MOE)
 - bad sampling methods (voluntary response, convenience, etc)
 - incomplete sampling frame (know what a sampling frame is)
 - Undercoverage (leaving out or not sufficiently taking into account part of a population)
 - Undercoverage - left out part of the population
 - doesn't take into account all parts of the population
 - makes the sampling frame incomplete
 - Sampling frame - a list of the population that we use for taking the sample
- Non Sampling errors - errors not resulting from the act of sampling
 - Processing errors
 - Response errors (Incorrect answers)
 - Non response errors
 - Response bias - answering the way they think the asker wants them to
 - Poorly worded questions - lead someone to an inaccurate response
- Stratified random samples
 - What it is and why we use it
 - Identify groups in the sample - groups are called strata
 - How to take it (stratify first, then random sample within strata)
 - Identify strata, and within each strata, take a random sample
- Cluster Sampling
 - What it is and why we use it
 - Identifying specific geographic areas, then we evaluate each subject within that area
 - How to take it
 - randomly sample cluster, then evaluate everything in cluster
- Response and Explanatory Variables
 - Know how to recognize and determine which is which
 - Response - what was the effect
 - how did the treatments affect the situation
- Subjects (like individuals)
 - the people that take part in the experiment
- Treatment
 - what is being done to subjects in an experiment
 - Know that treatments are different levels and/or different combinations of explanatory variables
- Lurking Variable
 - Definition
 - variables that are not part of our list of explanatory variables, but we are able to identify that they may have an impact on our study
 - Know how we attempt to control for (account for) lurking variables from comparing two or more treatments
- Confounded Variables
- Placebo Effect
 - takes place when somebody has received a treatment and they believe so firmly that there is gonna be an effect on them that they feel some kind of effect

- sugar pill
- pill that should not generate any kind of response
- Randomized Comparative Experiments
 - Compare two or more groups to control the effects of lurking variables
 - Randomization (hopefully) controls for bias/lurking variables
 - Large sample size control variability
 - the larger the sample, the less variable
- Statistical Significance
 - Definition
 - not likely to occur by chance
 - you are never 100%
 - How to interpret (like in homework)
 - We can say there is a statistically significant difference when comparing groups by comparing confidence intervals
 - Confidence intervals overlap \rightarrow NOT statistically significant
 - Confidence intervals do NOT overlap \rightarrow statistically significant
- How to live with observational studies
 - some situations where you can not do experiments
 - not ethically acceptable
 - anything that causes harm to subjects
- Double Blind neither the subject nor the person evaluating the study is aware of the treatment received
- Single Blind one side or the other does not know
- Problems with Real World Studies
 - Undercoverage
 - Refusal to Participate
 - Non adherers
 - do not follow the rules
 - can confound the outcomes of the study
 - Dropouts
 - people decide they don't want to do it anymore
 - may not be able to stay in experiment
- Goal is to generalize to the population
 - Statistical Significance
 - Realistic Setting
 - Repeatable Results
- Randomized Comparative Experiments
 - Compare two or more groups to control the effects of lurking variables
 - Randomizing to control bias
 - Using large sample sizes to control variability
 - We studied three designs
 - completely randomized design (analog to simple random sample)
 - all subjects randomly allocated to a treatment
 - Block design (analog to stratified random sample)
 - helps reduce variability in estimates
 - subjects divided into blocks based on similar characteristics
 - controls for lurking variables
 - Matched pairs
 - paired samples
 - special case where each pair is a block
 - only used when there are two treatments