

PROJECT 2: Solutions

Statistics 401: Fall 2006

Due Friday, September 15

Total: 30 pts.

1. (5 pts) A random sample of size $n = 500$ Montana taxpayers was obtained. Each member of the sample was asked whether or not the state should allow a partial income tax deduction for tuition expenses incurred by parents or guardians who send their children to private schools (including schools having religious affiliations).
 - (a) This is an observational study since no treatment is being administered.
 - (b) This is a simple random sample.
 - (c) Stratify Montana taxpayers into two strata. The first strata contains all Montana taxpayers with children. The second strata contains all Montana taxpayers without children. Take a simple random sample from each strata, where each sample size is proportional to the size of the strata in the population.
 - (d) This is a stratified random sample.

2. (9 pts) Forty psychotherapists and 212 people undergoing psychiatric care (clients) agreed to participate in a study on how therapist note-taking affects the relationship with a client. Twenty therapists and 106 clients were randomly assigned to each of two groups. The therapists in one group were told to take notes during each session with the clients in that group. The therapists in the other group were told to not take notes, but to write their case summaries after the client had departed. At the conclusion of the five sessions, a measure of the overall relationship of each therapist with his or her clients was taken.
 - (a) The population of interest is the therapists.
 - (b) The factor in this experiment is “note-taking,” which has two categories: “during session” and “after session.”
 - (c) The response in this study is the concluding psychiatric evaluation.
 - (d) The experimental design used in this study is called a completely randomized design or CRD.
 - (e) To deal with the extraneous variable “therapist experience,” one could block the therapists into two blocks. The first block contains therapists with high experience and the second block contains therapists with low experience. Therapists from each block will now be randomly assigned to “note-taking during a session” and “note-taking after a session” treatment groups.
 - (f) The new experimental design is called a Randomized Block Design.
 - (g) Some extraneous factors are food consumption; length, location, and time of day of the sessions; and attire of the therapist. One extraneous factor is the location of the therapy sessions. One can directly control for location by having all therapy sessions take place in similar parts of town in similar offices. One can directly control for the time of day of the therapy sessions by requiring that all sessions take place during the morning (before the therapist sees “regular paying” clients).

3. (2 pts) Random samples from a population are generally representative of the larger population regardless of the size of the population. We will see later that sample size depends on the variability of the population.
4. (2 pts) The conclusion depends on a sample which was not randomly chosen from the population of all California doctors. Rather, the sample is from 2000 California doctors who chose to complete and mail back the surveys. Thus, the sampling plan probably contains non-response bias. For example, the doctors who chose to respond could be more bitter than the others (and more motivated to respond) about managed care. So results of this survey can not be used to make conclusions about all California doctors.
5. (2 pts)
 - (a) It is unlikely that the researchers in the study randomly required some families to go get pets and required that the others not acquire pets (and to get rid of them if they had pets already). Thus, this study was probably an observational study.
 - (b) Perhaps people with pets tend to live in suburban and rural areas, and perhaps the location of raising a child is the true cause of lower allergy rates. Thus, family location is a potential confounding variable.
6. (6 pts)
 - (a) Attempt to collect a simple random sample from the MSU population. Due to non-responses, this is probably not possible. Randomly assign the individuals from the sample into one of two groups: one which listens to a Mozart piano sonata before taking some IQ test, and another which does not listen to Mozart before the test.
 - (b) We could directly control for time effects by requiring that all students listen to Mozart the same amount of time and that the test is taken by all individuals at the same time after the music was listened to. Other extraneous variables that can be directly controlled: music volume, temperature of the rooms where the test is administered, and the decor of the rooms where the test is administered to name a few.
 - (c) We might be interested in blocking by age, class-rank or gender. To control for the effect of gender, we could group the individuals into two blocks, one for males and one for females. Assign the males randomly to each of the two music groups. And assign the females randomly to each of the two music groups.
 - (d) Although it is unlikely that we can get a SRS from the population of MSU students, we must randomly assign the individuals to the treatment groups. This random assignment would allow us to make conclusions about the effect of listening to Mozart on IQ tests.
7. (1 pt) If one is interested in making an inference from a sample to a population, then random sampling is most important.
8. (1 pt) If one wishes to make a causal inference then random assignment is most important.
9. (1 pt) A cluster random sample is more efficient than a simple random sample when the population is “naturally” divided into roughly equally sized groups, each of which is representative of the total population.

10. (1 pt) A systematic random sample is more efficient than a simple random sample if the population list is either not related to the response variable, or if it is related, the relationship is not cyclic.