Name: _____ Date: _____
Math 455- Mrs. Bagala Practice

Sampling and Experimental Design Review

1. In each of the following examples **identify the sampling method** being used and **identify a potential source of bias**.

(a) A sample of 2,000 students was sought to estimate the average achievement in science of fifth graders in a city's public schools. The average fifth grade enrollment in the city's elementary schools is 100 students. Thus, 20 schools were randomly selected and within each of those schools all fifth graders were tested.

This is a cluster random sample with the elementary schools as heterogenous clusters. A potential source of bias could be the geographic location of the schools chosen. If most or all of the schools chosen end up being similar in terms of geographic location or socio-economic status, the sample may not be representative of the population.

(b) A researcher is interested in interviewing maximum-security inmates. She groups inmates by the type of crime committed and takes a simple random sample of inmates from each group.

This is a stratified random sample because the inmates were grouped into homogenous strata based on type of crime. A potential source of bias could be response bias as the inmates may not be honest about their crime.

(c) A researcher has a population of 100 third grade children from a local school district from which a sample of 25 children is to be selected. Each child's name is put on a list, and each child is assigned a number from 1 to 100. Then the numbers 1 to 100 are written on separate pieces of paper and shuffled. Finally, the researcher picks 25 slips of paper and the numbers on the paper determine the 25 participants.

This is a simple random sample as every individual and every group of size n has the same probability of being selected. A potential source of bias could be that 25 male students are selected and no female students are selected or vice versa.

(d) A researcher is interested in political attitudes of individuals within a particular state. The researcher uses the following procedure to obtain a sample: (1) a random sample of counties is drawn, (2) within the counties selected, districts are randomly drawn, (3) within each district, streets are randomly sampled, (4) within selected streets, building are randomly drawn, and (5) all residents within the buildings selected comprise the sample.

This is a multistage design as many different sampling methods are used. One potential source of bias could be that buildings chosen could be similar in terms of socio-economic status and will therefore not represent the entire population.

(e) Instructors teaching research methods are interested in knowing what study techniques their students are utilizing. Rather than assessing all students, the researchers ask every 8th student that enters the lecture hall for the final exam.

This is a systematic random sample. A potential bias could be response bias if students lie to the instructors about their research methods in fear of impacting their grade.

2. A survey will be conducted to examine the educational level of adult heads of households in the United States. Each respondent in the survey will be placed into one of the following categories:

- Does not have a high school diploma
- Does have a high school diploma

The survey will be conducted using a telephone interview. Random-digit dialing will be used to select the sample.

(a) For this survey, state one potential source of bias and describe how it might affect the estimate of the proportion of adult heads of households in the United States who do not have a high school diploma.

One potential source of bias in this sample design is undercoverage bias. This design method systematically ignores those that do not have a telephone. It is possible that those with lower paying jobs do not telephones and would therefore be ignored using this sample design. As a result of the potential undercoverage bias, the estimate of the proportion of adult heads of households in the US who do not have a high school diploma would be underestimated.

(b) Since education is largely the responsibility of each state, the agency wants to be sure that estimates are available for each state as well as for the nation. Identify a sampling method that will achieve this additional goal and briefly describe a way to select the survey sample using this method.

A stratified random sampling technique could be employed to ensure that there are samples from each state. The researchers could use each state as a strata and selected simple random samples of indiviuals from each state to make up the sample. After stratifying by state, researchers could use a list of residents to select a simple random sample from each state.

3. A local school board plans to conduct a survey of parents' opinions about year-round schooling in elementary schools. The school board obtains a list of all families in the district with at least one child in an elementary school and sends the survey to a random sample of 500 families. The survey question is provided below.

A proposal has been submitted that would require students in elementary schools to attend school on a year-round basis. Do you support this proposal? (Yes or No)

The school board received responses from 98 of the families, with 76 of the responses indicating support for year-round schools. Based on this outcome, the local school board concludes that most of the families with at least one child in elementary school prefer year-round schooling.

(a) What is a possible consequence of nonresponse bias for interpreting the results of this survey?

Only 98 out of 500 families or 19.6% responded to the survey. It is likely that the opinions of parents who are most in favor of year-round schools were the ones who responded and there are over-represented in the survey.

(b) Someone advised the local school board to take an additional random sample of 500 families and to use the combined results to make their decision. Would this be a suitable solution to the issue raised in part (a)?

No, this would not be reasonable. Even after increasing the sample size the non-response bias would still be present because there are still 402 of 500 families that did not respond from the first survey.

(c) Suggest a different follow-up step from the one suggested in part (b) that the local school board could take to address the issue raised in part (a).

The school board could contact the 402 families who did not respond and try to get their responses. Alternatively, the school board could take a completely new survey using phone calls or person to person contact to try to eliminate the amount of non-response bias.

4. Turkeys raised commercially for food are often fed the antibiotic salinomycin to prevent infections from spreading among the birds. However, salinomycin can damage the birds' internal organs, especially the pancreas. A researcher believes that a combination of selenium and vitamin E in the birds' diet may prevent injury. He wants to explore the effects of two different dosages of selenium (call them S1, S2) in combination with any of three different dosages of vitamin E (call them E1, E2, E3) added to the turkeys' diets. There are 48 turkeys available for the study. At the end of the study, the condition of the bird's pancreas will be examined.

(a) How many treatments are there in this experiment? Identify all factors and levels.

(b) Outline in diagram form an appropriate design for this experiment. In your diagram, indicate how many turkeys are assigned to each treatment group.

(c) Describe how randomization was used in the experiment. Why is randomization important?

5. Because of concerns about employee stress, a large company is conducting a study to compare two programs (tai chi or yoga) that may help employees reduce their stress levels. Tai chi is a 1,200 year-old practice, originating in China, that consists of slow, fluid movements. Yoga is a practice, originating in India, that consists of breathing exercises and movements designed to stretch and relax muscles. The company has assembled a group of volunteer employees to participate in the study during the first half of their lunch hour each day for a 10 week period. Each volunteer will be assigned at random to one of the two programs. Volunteers will have their stress levels measured just before beginning the program and 10 weeks later at the completion of it.

(a) A group of volunteers who work together in the same department ask to be assigned to the same program so that they can participate in that program together. Give an example of a problem that might arise if this is permitted. Explain to this volunteer group why random assignment to the two programs will address this problem.

For example, a deadline in the department where the group of volunteers works has been moved back, lowering the stress levels of those working in the department. If the volunteers from this department were all in the same treatment group, this change in stress level could mistakenly be attributed to the treatment.

Without random assignment of volunteers to the two programs, it is possible that the two treatment groups could differ in some way that affects the outcome of the experiment. Randomization "evens out" the possible effects of potentially confounding variables.

(b) Someone proposes that a control group be included in the design as well. The stress level would be measured for each volunteer assigned to the control group at the start of the study and again 10 weeks later. What additional information, if any, would this provide about the effectiveness of the two programs?

Without the control group, the company could compare the two treatments, but would not be able to say whether the observed reduction in stress was attributable to participation in the programs. For example, a change in the work environment during this period might have reduced the stress level of all employees. The addition of a control group would enable the company to assess the magnitude of the mean reduction attributable to each treatment, as opposed to just determining if the two programs differ.