Learning Guide: Basics of Sampling
REVISED August 2019

I know that there are a lot of questions here. Sampling is complicated, important, and hard to understand. I apologize for the extensive list of questions, but I start with the assumption that this is your introduction to all concepts in sampling. Some of these questions will be easy for those of you who have previous knowledge and experience, but I would rather include them than cause people to miss important concepts. I have not included the normal list of questions you should be able to answer after class in this guide because we will spend time clarifying confusing ideas or concepts that you raise and discussing. You should be able to answer these questions after completing the required materials (readings and videos).

1. What is the difference between a theoretical population, an accessible population, and a sample?

2. What is the difference between a sample (of cases) and a census?

3. Why is a census preferable to a sample, *assuming you have the time and money to conduct a census*?

4. Students sometimes use the phrase “sample population” and this phrase occasionally appears in the literature. Why is it an erroneous phrase?

5. Researchers normally select a sample from the accessible population rather than the theoretical population. What must the researcher do to establish that the accessible population is a valid choice?

6. Explain the key differences between probability and non-probability samples in your own words.

7. Why are probability samples highly desirable (usually necessary) when you want to generalize specific traits of a sample to an accessible or a theoretical population – especially quantified traits like mean SAT scores of entering UF first-year students?

8. Three main factors affect your ability to extend conclusions to people or thing you did not study: non-sampling error (also called experimental error), sampling error (or sample variance), and sample bias.

    a. What is non-sampling or experimental error? What can you do to reduce this threat to your ability to generalize conclusions?

    b. What is sampling error or sample variance? What can you do to reduce this threat to your ability to generalize conclusions? Is it possible to eliminate completely this threat to generalization? Why or why not?

    c. What is sample bias? What can you do to reduce this threat to your ability to generalize conclusions? Is it possible to eliminate completely this threat to generalization? Why or why not?

9. Are the respondents and the sample necessarily the same people (or things)? Why or why not?

10. What is the response rate for a study? How to you determine it (VERY simple math)?

11. Which of the following constitute “non-response”?

    a. You are unable to contact someone who was randomly selected
    b. Someone declines to participate
    c. Someone refuses to sign the informed consent statement
    d. Someone starts to complete your on-line **questionnaire** (note, not survey), but stops answering questions at question number 13 out of 20 questions
    e. Someone starts to complete your on-line **questionnaire**, but stops answering questions at question number 7 out of 20 questions

12. Why is non-response (nonresponse bias) a threat to your ability to generalize the conclusions you reach in a research project? Why doesn’t replacement solve remove the threat?

13. What is the difference between coverage bias and selection bias? Assume you want to conduct a study to examine the impact of participation youth groups engaged in community service during high school on academic achievement in four-year colleges and universities. Think about coverage bias and EPSEM (equal probability selection method) as you answer these questions.

    a. Would UF be a good place to conduct this study? Why or why not?
    b. Assume you decide to conduct the study with UF undergraduate students. UF student records indicate participation in youth organizations. The registrar has offered to provide a list of randomly selected e-mail addresses of students who reported that they belonged to a youth organization in high school. Which would be better – to use the list of students offered by the registrar’s office or to ask the registrar to send a randomly selected list of UF students as a whole?

14. I ask you **not to use the term survey** for a number of reasons. One reason is that people substitute the word “survey” for “sample.” (They also confuse it with a method of data collection and a type of research design.) Results of the failure to use clearly defined terms include erroneous results, unjustified conclusions, and poor practice. Here are some statements made by a company purporting to conduct on-line surveys of the “general public.” Identify the errors with regard to sampling these statements reveal.

    “The average response rate for email surveys is 24.8%. In a survey of the general public a response rate of 24.8% is good compared to the telephone standard of 8-12%. It could be desirable to increase this number, but there is no reason to panic if you are have response rates lower than 25%. You may have a difficult to reach sample group or your survey topic may be sensitive. In any case, your response rate will differ based on the topic and target audience, no matter how good your research plan and even the best surveys can be tweaked to increase response rate.”

15. What is the difference between a random sample and a probability sample? Are all random samples probability samples? Are all probability samples random samples?

16. It is true that strictly speaking, if even one person you selected for a sample refuses to participate, you no longer have a probability sample. Why?

17. In reality, we live with less than perfect probability samples all the time and in fact most samples are really random samples, NOT true probability samples. The reason is very simple – it is extremely difficult to get a true probability sample, not just in the social sciences but also in many other sciences like biology, geology, climatology, and yes, physics. When is a random sample "good enough" to convince you that the **conclusions**, not the specific numbers, but the theoretical conclusions, drawn in a study can be generalized? What will you look for in descriptions of sampling procedures to convince you that an author can produce valid (warranted) conclusions ***and*** generalize those **conclusions** to the population of interest?

18. What is effect size?

19. What is the difference between effect size and significance level?

20. What role does effect size play in determining (1) sample size and (2) generalizing statistical results and general conclusions?

21. Which of the three threats to generalization does effect size address – non-sampling error, sampling error (variance), or bias?

22. How do you know what effect size you need for a result to be both statistically significant ***and*** meaningful?

23. Should you even worry about effect size if you are not taking a random sample?