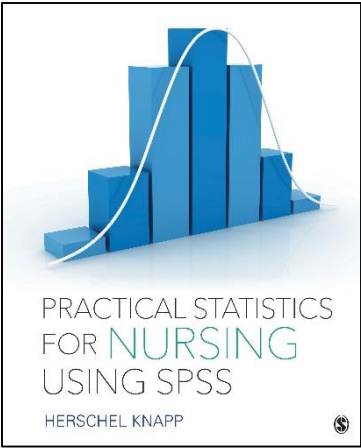


Chapter 2
Sampling
Solutions to All Exercises



Contents

Exercise	Page
2.1	2
2.2	3
2.3	4
2.4	5
2.5	6
2.6	7
2.7	8
2.8	9
2.9	10
2.10	11

Exercise 2.1

(a)

In terms of time, gathering a sample of an entire population could take weeks, months, or possibly years depending on the size of the population, geographical area involved, and complexity of the data to be gathered. The point of research is to get the answer to a meaningful question, presumably to solve a problem or improve a situation. Such answers can be expedited by sampling, as opposed to gathering data on the entire population. Attempting to sample the entire population would be so time-consuming that by the time you had acquired the full data set, the nature of the initial problem may have changed substantially, thereby making your findings virtually irrelevant.

(b)

No matter what method of gathering data one uses (e.g., in-person surveys, phone surveys, postal/email contact, experimental designs, tests, etc.), costs are involved. When it comes to data collection, subject participation fees are customary, data collection teams need to be paid, and there are administrative costs involved (photocopying, office space, equipment, office supplies, etc.). Considering the volume of individuals that would be involved in studying an entire population, the costs would be prohibitive.

(c)

Beyond the lengthy time and exorbitant costs that would be involved in studying an entire population, it is seldom feasible to gather data on so many people. The population may span a broad geographical region, requiring lengthy travel among the research staff. Within a population, one would likely encounter a variety of languages; it may not be possible to translate verbal or written instructions to attain a full population sample. In addition, some studies may involve experiments that can be administered only at a special facility; it would be impossible to arrange a round-trip commute for every member of a population to that facility to participate.

(d)

Extrapolation involves working diligently to gather a representative sample in sufficient quantities to facilitate stable statistical processing. If the data gathered constitutes a representative sample, then one could (more) confidently extrapolate the findings to (better) comprehend the overall population from which it was drawn; this is also known as “external validity”.

Exercise 2.2

(a)

The population is the total census; basically, it includes every single person within a specified domain. An example of a population would be every single patient in the hospital.

(b)

The sample frame is a smaller list derived from the population consisting of the individuals or records that could be actually accessed by the researcher. An example of a sample frame would be the list of patients who have been discharged from the hospital and who have given permission for the hospital to contact them to conduct quality of care surveys.

(c)

The sample is derived from selections made from the sample frame. For example, if the sample frame has 100 names and I only had the time (or budget) to sample 20 of them, then the sample would consist of the 20 that I'd randomly select from the 100 in the sample frame.

(d)

A representative sample has the same characteristics as the overall population that it was drawn from. A representative sample helps to facilitate external validity – meaning that what I learn from my sample can be plausibly generalized to describe the overall population that it was drawn from. For example, suppose the population contains 1,000,000 people consisting of 60% females, and 40% males. To gather a representative sample, I should have 60% females and 40% males, even if my sample is as low as 100 (I should have 60 females and 40 males).

Exercise 2.3

(a)

The population is the 20,000 (unique) patients who have accessed this clinic.

(b)

The sample frame is the list of 1,000 patients who have appointments this week (200 patients per day x 5 days = 1,000).

(c)

Since I want to randomly gather data from 10% of the patients, I would get 10 coins and write an X on one of them. I'd put the coins into a hat, and without looking, I'd randomly select 1 coin each time a patient entered the clinic. If I draw the coin with the X on it, then I'd survey that patient. After each coin draw, I'd drop the coin back into the hat, hence, the likelihood of being selected (1 out of 10 chance) would remain the same throughout the process.

(d)

I would approach the indicated patients, 1 at a time and ask if they'd be willing to answer a survey that takes less than a minute. I would have a clipboard with a 3 column form on it where I'd record: (1) Patient's home address, (2) Mode of transportation, (3) Round trip commute time.

Exercise 2.4

(a)

The population is the 250 (unique) patients who go to this dialysis center.

(b)

The sample frame is the 120 patients who have consented to being contacted.

(c)

Since this is systemic sampling, I'd need to derive the skip term (k), which would be the sample frame divided by the target sample size ($120 \div 20 = 6$). Next, I would need to get the start point, which would be a random number between 1 and k (6). For this example, let's say the start point is 2. I would then be able to derive the sample numbers (starting with 2 and then adding 6 again and again, until I had the sample of 20): 2, 8, 14, 20, 26, 32, 38, 44, 50, 56, 62, 68, 74, 80, 86, 92, 98, 104, 110, 116.

(d)

I would use the sequence of (20) numbers to identify the 2nd, 8th, 14th, etc. patients listed on the call-sheet and contact them in that order to administer the survey.

Exercise 2.5

(a)

The population is the 2,500 students enrolled at the high school.

(b)

The sample frame is the 2,400 students who are on campus that day (2,500 enrolled – 100 absent).

(c)

Presuming that each student has (exactly) one P.E. (Physical Education) class, before we went out to the school, we would coordinate with the school administrator and P.E. department and ask them if I could conduct BMI screenings among their students using a special hand-held device. On the day of the testing, we'd set up our test stations in the gym and explain to the students that their participation is voluntary, and that we'll be gathering their BMI (using a device that is harmless and that doesn't hurt) and asking them their age. We would then provide a briefing to the students demonstrating how to properly grip the BMI device. Next, we'd then ask the students to line up at each table. We might instruct the girls to line up at one table and the boys to line up at another table – that would help keep our data collection more gender balanced.

After they've been tested, they'd be dismissed to the playground for their regular P.E. class and not admitted back into the gym – that way, we can avoid double-counting.

(d)

To gather the data, I would use a clipboard with a prepared 3 column form: (1) BMI, (2) Gender F / M, (3) Age.

Exercise 2.6

(a)

The population is the 6,000 households in Cityville.

(b)

I would divide the target sample size by the number of blocks ($900 \div 300 = 3$); this would tell me that I'll be sampling 3 households per block (for the 300 blocks). Next, I would go to the list of addresses for each block, and randomly select 3 addresses from each block.

(c)

I would distribute unique lists of addresses to the members of my survey team and have them go out to gather the data.

Exercise 2.7

(a)

I would want to gather a relatively representative sample, so I would want to gather data from people entering a local grocery store (presuming most everyone in the community goes grocery shopping). I'd confer with the manager and ask permission to briefly talk to people outside the store; I'd also show the data collection sheet along with a briefing detailing the study.

(b)

I would introduce myself and state that I'm part of a health science research team investigating smoking in this community. I would then ask them their first name, just to be friendly, and then I'd ask if the person smokes and, if so, about how many cigarettes per day. I'd record their responses on a clipboard. Whether they choose to respond or not, I'd thank them.

Exercise 2.8

(a)

To get things started, I would ask the tutors if they could refer me to any students who had dyslexia. If they provided any referrals, I would begin by interviewing the students that they suggested. If the tutors didn't have any names of students, then I would discretely ask each student if they were dyslexic; if they respond "yes", then I would ask them if they'd be willing to respond to the survey.

(b)

Upon concluding each survey, I'd ask the person if they know of any other people who have dyslexia who might be willing to take part in the survey. I would also ask if they knew of any other tutoring or educational centers, or other places where I might find some more people with dyslexia.

Exercise 2.9

(a)

I would stand in the lobby of the Fitness Center starting at 8:00 A.M. (when they open). At first I'd approach anyone and tell them that I'm conducting a brief survey of the gym members. If they agree to talk to me, I would begin by asking them their age. Once I'd hit my quota for a group (25 minors, 50 adults), I'd stop approaching that group and proceed with the other group. When I'd gathered all 75 surveys, I'd stop the sampling process completely and move on to the analysis.

(b)

I would approach potential participants, identify myself as a staff member of the Acme Fitness Center (I'd point to my badge), and ask the person if I could ask them a few questions. I would record their responses on a clipboard, which would contain the printed survey. The first question I would ask is "What is your age?" which I would record on the survey form; if they didn't want to tell me their age, I would ask "Are you 18 or over?" unless it was obvious that the person was clearly either a child or an adult. This way I could keep a running count of how many minors and how many adults I had surveyed. If the quota was full (e.g., I already had surveyed 50 adults) and a person replied that their age is over 17, then I would thank them for their willingness to take the survey, but explain that we're only gathering data from minors right now.

Exercise 2.10

(a)

I'd put up flyers in nearby neighborhoods to advertise this cooking class. The flyer would indicate the dates and times for these classes, the eligible ages for participation (10 – 14 years old), and that parental consent is required. I would provide tear-tags on the bottom of the flyer so anyone who is interested could call the phone number to reserve a spot in the class. During that call, we'd ask to speak with the parent or guardian and tell them about the healthy cooking course, and explain that we'd also be gathering height, weight and BMI data from each participant. Even though these are non-invasive procedures, they we'll need a consent form signed by them (which we could email or postal mail to them).

(b)

On the first day of the class, I would individually gather data on each student (height, weight using a medical scale) and use the data to calculate their BMI. The information would be written on prepared forms, which would be used to collect the data again at the end of the classes.