1

00:00:08,370 --> 00:00:10,600

Welcome to Step 2.

2

00:00:10,600 --> 00:00:14,870

You have just finished Step 1 and now you

are ready for Step 2.

3

00:00:14,870 --> 00:00:20,470

We will start by talking about sampling designs

and sampling frames—that is, how to select

4

00:00:20,470 --> 00:00:22,980

the households you will survey.

5

00:00:22,980 --> 00:00:28,529

In the next video, we will get into how to

choose sample sizes.

6

00:00:28,529 --> 00:00:32,960

Suppose we want to know the child labor prevalence

in our population of interest.

7

00:00:32,960 --> 00:00:40,180

But it is too expensive to survey all households,

so we survey a sample, shown here in blue.

8

00:00:40,180 --> 00:00:45,250

We can calculate the prevalence of child labor

in this sample of households and use that

9

00:00:45,250 --> 00:00:50,859

number to estimate the prevalence of child

labor in our population of interest.

10

00:00:50,859 --> 00:00:52,660

How many households to select?

11

00:00:52,660 --> 00:00:55,489

And how do we select these households?

12

00:00:55,489 --> 00:01:00,899

A sampling design includes a sample size,

which is the number of households we select,

13

00:01:00,899 --> 00:01:05,850

and a sample structure, which defines the

way we select these households.

14

00:01:05,850 --> 00:01:10,500

The bigger the sample size the more precise

our estimate will be and the more accurately

15

00:01:10,500 --> 00:01:14,890

we will learn about the child labor prevalence

in the population.

16

00:01:14,890 --> 00:01:19,920

The standard error measures how close we expect

our estimate to be to the true child labor

17

00:01:19,920 --> 00:01:25,400

prevalence in the population. It is sometimes called the sampling error.

18

00:01:25,400 --> 00:01:28,480

Smaller sample sizes give larger standard

errors.

19

00:01:28,480 --> 00:01:32,910

In other words, with smaller samples, we are

less certain about the child labor prevalence

20

00:01:32,910 --> 00:01:35,000

in the population.

21

00:01:35,000 --> 00:01:39,530

Here is a different sample structure, where

we select households in groups.

22

00:01:39,530 --> 00:01:42,720

The sample structure also affects the standard

error.

23

00:01:42,720 --> 00:01:45,720

We will talk more about this soon.

24

00:01:45,720 --> 00:01:50,640

We want all households to have the same chance

to be in the sample, so that the sample looks

25

00:01:50,640 --> 00:01:52,800

like the population.

26

00:01:52,800 --> 00:01:56,260

These are called “equal probability” sampling

designs.

27

00:01:56,260 --> 00:02:01,770

The simplest equal probability design is a

simple random sample, called “SRS,” which

28

00:02:01,770 --> 00:02:05,080

allows any sets of households to be sampled.

29

00:02:05,080 --> 00:02:09,950

However, non-SRS designs can also be equal

probability.

30

00:02:09,950 --> 00:02:12,080

Here are households sampled by SRS.

31

00:02:12,080 --> 00:02:16,340

SRS sounds appealing because it is so simple.

32

00:02:16,340 --> 00:02:21,799

But this design requires that you have a list

of all households in the population of interest.

33

00:02:21,799 --> 00:02:24,090

This list is called a sampling frame.

34

00:02:24,090 --> 00:02:28,959

Usually, recent lists of all households are

not available and would take a long time to

35

00:02:28,959 --> 00:02:32,659

create for the entire area covered by the

project.

36

00:02:32,659 --> 00:02:37,689

Instead, suppose all that is available is

a list of the 21 census enumeration areas

37

00:02:37,689 --> 00:02:40,209

in our population of interest.

38

00:02:40,209 --> 00:02:44,060

We can sample a few enumeration areas from

this list first.

39

00:02:44,060 --> 00:02:49,450

Then, we would only have to create a list

of households in those areas.

40

00:02:49,450 --> 00:02:52,709

This saves a lot of time and money.

41

00:02:52,709 --> 00:02:55,980

This method is called cluster sampling.

42

00:02:55,980 --> 00:03:00,379

In our example, enumeration areas are the

clusters.

43

00:03:00,379 --> 00:03:05,920

They are also called primary sampling units,

or PSUs, because they are the first things

44

00:03:05,920 --> 00:03:07,650

we sample.

45

00:03:07,650 --> 00:03:11,310

The first stage of sampling is to sample the

clusters.

46

00:03:11,310 --> 00:03:16,010

Then we list all the households in those clusters.

47

00:03:16,010 --> 00:03:21,840

The second stage of sampling is to sample

some households from the sampled clusters.

48

00:03:21,840 --> 00:03:26,790

Notice that the enumerators will only have

to travel to the sampled clusters, reducing

49

00:03:26,790 --> 00:03:28,370

travel.

50

00:03:28,370 --> 00:03:33,760

This basic sampling design is called two-stage cluster sampling.

51

00:03:33,760 --> 00:03:39,580

It is not a simple random sample, or SRS,

because we select clusters before we select

52

00:03:39,580 --> 00:03:41,079

households.

53

00:03:41,079 --> 00:03:46,360

But it can be equal probability, meaning that

all households have the same chance of being

54

00:03:46,360 --> 00:03:49,140

selected for the sample.

55

00:03:49,140 --> 00:03:55,480

The sample size template will help you choose

sample sizes to make this design equal probability.

56

00:03:55,560 --> 00:03:59,160

We will show you how to do this in the next video.

57

00:03:59,280 --> 00:04:05,019

Let’s go back to Stage I, where we sampled

clusters, or PSUs.

58

00:04:05,019 --> 00:04:08,860

What if we got unlucky and selected these

clusters?

59

00:04:08,860 --> 00:04:13,190

This would be bad, because our sample does

not look like the population.

60

00:04:13,190 --> 00:04:18,540

We are missing households from District C

and have too many from District A.

61

00:04:18,540 --> 00:04:23,830

Instead, we might want to take separate samples

in each district.

62

00:04:23,830 --> 00:04:26,560

This is called stratified sampling.

63

00:04:26,560 --> 00:04:30,090

In our example, districts are the strata.

64

00:04:30,090 --> 00:04:35,530

Other examples of strata include regions or

urban versus rural.

65

00:04:35,530 --> 00:04:40,729

The sample size template, which we will discuss

in the next video, will help you choose sample

66

00:04:40,729 --> 00:04:45,949

sizes in each stratum to make this design

equal probability.

67

00:04:45,949 --> 00:04:51,340

Stratified sampling is useful to help the

sample look more like the population in terms

68

00:04:51,340 --> 00:04:54,370

of categories like districts.

69

00:04:54,370 --> 00:04:58,590

We might also want the sample to look like

the population in terms of something like

70

00:04:58,590 --> 00:05:01,940

PSU size or household size.

71

00:05:01,940 --> 00:05:05,600

Here is our population of households ordered

by size.

72

00:05:05,600 --> 00:05:10,889

With simple random sampling, we could get

unlucky and select only the small households,

73

00:05:10,889 --> 00:05:12,960

as we see here.

74

00:05:12,960 --> 00:05:21,300

Instead, we can order the households by size

and take a sample that is spread across all sizes.

75

00:05:21,419 --> 00:05:24,860

Now open up your Sampling Worksheet to Step

2.

76

00:05:24,860 --> 00:05:30,050

The first question is whether you plan to

use the basic two-stage sampling design.

77

00:05:30,050 --> 00:05:35,000

We think you probably do—it will make your

job a lot easier!