Systematic Sampling

Lecture 9A
Section 2.7

Robb T. Koether

Hampden-Sydney College

Wed, Sep 10, 2008

Outline

Exercise 2.23, p. 116

- A class consists of 100 students. Suppose that we are interested in the heights of the people in the class. We could take a stratified random sample, using gender as the stratification variable.
- Suppose that it is known that 60 of the 100 students in the class are female.
- Someone shakes a box containing the 60 female tags, and with a blindfold on, reaches in and draws 10 tags.
- Then, this same person shakes a box containing the 40 male tags, and with a blindfold on, reaches in and draws 10 tags.

Exercise 2.23, p. 116

- Now we have two simple random samples, one for the males and one for the females, each of size 10.
- Suppose that the average height of the males in the sample is 70 inches, while the average height for the females in the sample is 63 inches.
- How should these two average heights be combined to produce an estimate of the average height of the people in the class?

Solution

- We should use a weighted average.
- Weight the female average with 0.60 and the male average with 0.40.
- The estimated average for the class is

$$\begin{aligned} \text{est. avg.} &= (0.60)(63) + (0.40)(70) \\ &= 37.8 + 28 \\ &= 65.8 \end{aligned}$$

Systematic Sampling

Definition (1-in-k systematic sampling)

1-in-k systematic sampling is a sampling method in which one of the first k members of the population is selected at random. Then beginning with that member, every k^{th} member is selected.

- For example, if $k = 10$, then one of the first 10 members is selected at random. Suppose member #6 is selected. Then beginning with member #6, every 10^{th} member is selected, i.e., members 6, 16, 26, 36, and so on.

Systematic Sampling

- To select a 1-in-k systematic sample,
  - Number the members of the population 1 through N.
  - Using `randInt(1,k)`, choose a random starting point from 1 to k. That represents the first member of the sample.
  - From that starting point, put every k^{th} member in the sample.

Example

- Let's choose a 1-in-5 systematic sample from the 22 students in this class.
- Use `randInt(1,5)` to choose the first member.
- Then choose every 5^{th} student thereafter.

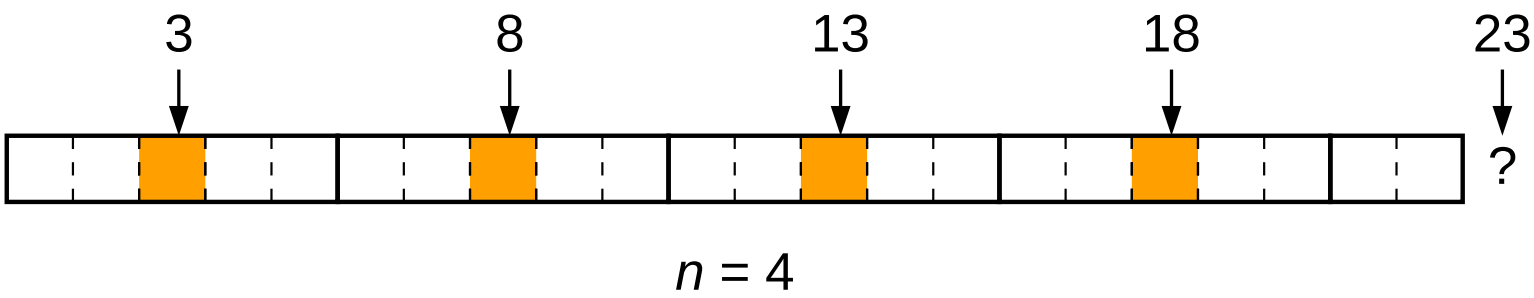
The Sample Size

- In the last example, notice that $22/5 = 4.4$.
- So the sample size could end up being either 4 or 5, depending on the starting point.
- If the starting point is 3, then we get 4 members (3, 8, 13, 18).
- If the starting point is 2, then we get 5 members (2, 7, 12, 17, 22).
- Either way, each member has a $1/5$ chance of being in the sample.

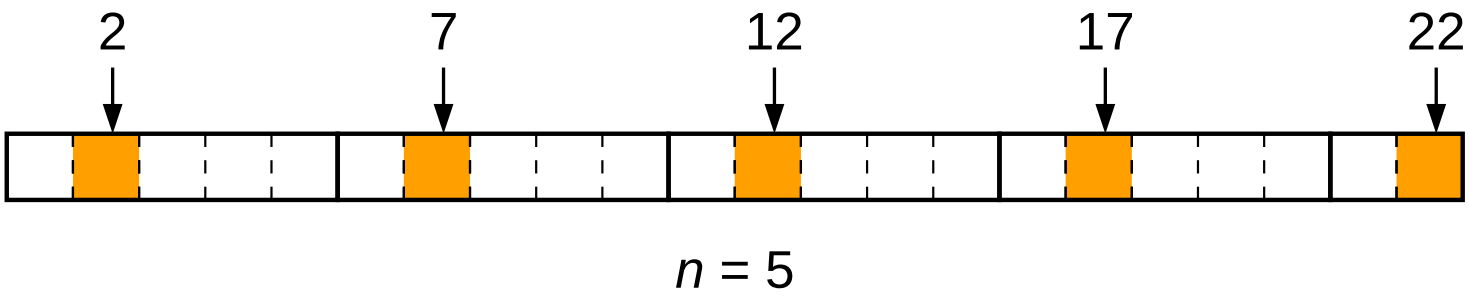
The Sample Size

$N = 22$

The Sample Size



The Sample Size



How to Determine k

- To determine k,
  - Divide the population size N by the desired sample size n.
  - Let $k = N/n$ (rounded off).

Example

- Suppose we want to select $n = 12$ members from a population of size $N = 187$.
- Compute $k = 187/12 \approx 16$.
- Use `randInt(1,16)` to get a random integer from 1 to 16. Say we get 9.
- Then beginning with member #9, select every 16^{th} member:

  9, 25, 41, 57, 73, 89, 105, 121, 137, 153, 169, and 185.

The Language of Studies

Lecture 9B

Sections 3.1 - 3.3

Robb T. Koether

Hampden-Sydney College

Wed, Sep 10, 2008

Two Types of Variable

Definition (Response variable)

A response variable is a variable whose value is influenced by other variables in the study.

Definition (Explanatory variable)

An explanatory variable is a variable whose value influences other variables in the study.

Example

- In a study, a standard drug and an experimental drug are administered to patients with fevers in order to reduce the fevers.
- The explanatory variable is the drug which was administered to the patient.
- The response variable is whether the patient's fever was reduced.

Observational and Experimental Studies

Definition (Observational study)

An observational study is a study in which none of the explanatory variables are manipulated.

Definition (Experimental study)

An experimental study is a study in which at least one of the explanatory variables is manipulated.

Observational or Experimental

- If an experimental study gives the researchers more control over the explanatory variables, then why would anyone conduct an observational study?

Observational or Experimental

- Suppose researchers wish to determine whether there is a relationship between drunk driving and traffic fatalities.
- Should this be an observational or an experimental study?
- Why?

Observational or Experimental

- A traffic engineer is studying the traffic flow at a particular intersection.
- He needs to know the average number of cars that turn left, turn right, and go straight at various times of the day.
- Should this be an observational or an experimental study?
- Why?

Levels and Treatments

Definition (Level)

A level is a value of an explanatory variable.

Definition (Treatment)

A treatment is a combination of values (levels) of two or more explanatory variables.

A Third Type of Variable

Definition (Confounding variable)

A confounding variable is a variable whose effect on the response variables cannot be separated from the effect of the explanatory variables.

- If a study has one or more confounding variables, then the researchers cannot attribute changes in the response variables to any one explanatory variable.

Do “Explanatory" Variables Really Explain?

- A study cannot *prove* that variations in the explanatory variable really are the cause of variations in the response variable.
- The study can only give evidence supporting that belief.
- It may be the case that there is a third variable that is affecting both the explanatory and response variables.
- It may be conceivable that the “response" variable affected the “explanatory" variable!

Do “Explanatory" Variables Really Explain?

- In the following situation, could there be any confounding variables? What?
  - The higher a student’s family income, the better his grades in school.
- In the following situation, is it clear which is the explanatory variable and which is the response variable?
  - The more someone drinks (alcohol), the more likely he is to get bad grades.

Evidence of Causation

- The following give evidence, but not proof, of causation.
  - The same association between the explanatory and response variables occurs in a variety of situations.
  - There is a plausible explanation of how the explanatory variable could affect the response variable.
  - There is no equally plausible third factor that could be affecting both the explanatory and the response variables.

Assignment

Homework

- Read Sections 2.7, 3.1 - 3.3, pages 117 - 121, 145 - 155.
- Let's Do It! 2.9, 2.10, 3.1, 3.2, 3.3.
- Page 121, exercises 28, 30 - 32.
- Page 155, exercises 1 - 3, 5 - 8.
- Chapter 2 review, p. 136, exercises 44, 46, 48 - 50, 51, 53, 55, 56, 63.