

Worksheet 5

Author:

Discussants:

=====

The following worksheet will give you practice understanding sampling bias and a little practice with for loops and constructing sampling distributions. Please answers the following questions and submit a compiled pdf with your answers to Moodle by 11:59pm on Sunday October 14th.

A list of functions we have used in class can be found on a shared class google document here

Part 1: A few Lock 5 exercises examining sampling bias

Lock5 exercise 1.54 (1st edition): Teaching Ability In a sample survey of professors at the University of Nebraska, 94% of them described themselves as “above average” teachers. Please answer the following questions:

- a) What is the sample? What is the population?
- b) Based on the information provided, can we conclude that the study suffers from sampling bias?
- c) Is 74% a good estimate for the percentage of above-average teachers at the University of Nebraska? If not, why not?

Answers:

- a)
- b)
- c)

Lock5 exercise 1.56 (1st edition): Effects of Alcohol and Marijuana In 1986 the Federal Office of Road Safety in Australia conducted an experiment to assess the effects of alcohol and marijuana on mood and performance. Participants were volunteers who responded to advertisements for the study on two rock radio stations in Sydney. Each volunteer was given a randomly determined combination of two drugs, then tested and observed. is the sample likely representative of all Australians? Why or why not?

Answer:

Lock5 exercise 1.60 (1st edition): The Centers for Disease Control and Prevention (CDC) administers a large number of survey programs for monitoring the status of health and health care in the US. One of these programs is the National Health and Nutrition Examination Survey (NHANES), which interviews and examines a random sample of about 5000 people in the US each year. The survey includes questions about health, nutrition, and behavioral while the examination includes physical measurements and lab tests. Another program is the National Hospital Ambulatory Medical Care Survey (NHAMCS), which includes

information from hospital records for a random sample of individuals treated in hospital emergency rooms around the country. Please answer the following questions:

- a) To what population can we reasonably generalize findings from NHAMES?
- b) B) To what population can we reasonably generalize findings from NHAMCS?
- c) For each of the questions below, indicate which survey, NHANES or NHAMCS, would probably be more appropriate to address the issue.
 - i. Are overweight people more likely to develop diabetes?
 - ii. What proportion of emergency room visits in the US involve sports-related injuries?
 - iii. Is there a difference in the average waiting time to be seen by an emergency room physician between male and female patients?
 - iv. What proportion of US residents have visited an emergency room within the past year?

Answers:

- a)
- b)
- c)
- i.
- ii.
- iii.
- iv.

Part 2: Practice with for loops and creating sampling distributions in R

Exercise 2a: The following code prints the numbers from 1 to 5. Please modify it to print the numbers from 2 to 10.

```
for (i in 1:5) {  
  print(i)  
}
```

```
## [1] 1  
## [1] 2  
## [1] 3  
## [1] 4  
## [1] 5
```

Exercise 2b: Recall that if we have a vector called `my_vec`, could store the number 27 in the first entry of the vector using the code `my_vec[1] <- 27`. Please modify the code below to create a vector called `my_vec` that stores the all the even numbers from two to 100. Also print the vector to make sure that it has the numbers you expect.

```
my_vec <- NULL # initialize the vector to an empty vector so we can store things in it
for (i in 1:5) {

  my_vec[i] <- i^2

}

my_vec

## [1] 1 4 9 16 25
```

Exercise 2c In class we created a sampling distribution for the word lengths of the Gettysburg address. Please use a for loop to recreate this sampling distribution using a sample size of $n = 5$, and have 10,000 points in your sampling distribution. Then plot a histogram of this distribution and describe its shape below. Also describe what each point in this sampling distribution corresponds to.

```
load("/home/shared/intro_stats/cs206_data/gettysburg.Rda")
```

Answers:

Reflection

How did this worksheet go? Are you feeling ok with the concepts of sampling bias and sampling distributions and with using for loops?