Module 2 Assignment 1

Ellen Bledsoe

2022-10-11

Simple Random Sampling – Rocks!

Assignment Details

Purpose

The goal of this assignment is to use R to calculate and understand accuracy of estimates.

Task

Write R code to successfully answer each question below or write text to successfully answer the question.

Criteria for Success

- Code is within the provided code chunks
- Code chunks run without errors
- Code produces the correct result
- Text answers correctly address the question asked

Due Date

March 17 by 2pm MST

Assignment Questions

Short Answer

Think back to how we sampled the rock data. Answer the following question in terms of our data (aka, "big square," "little squares," "dark rocks," etc.)

1. Define the *sample* standard deviation of the mean in terms of our data (not the equation but what we use to calculate standard deviation, what it tells us, and what it is an estimate of).

   Answer: the sample standard deviation is a measure of how much variation there is in the number of rocks counted in each little square. The sample sd is calculated from the 8 sample units (little squares) that we counted and is an estimate of *sigma*, the population standard deviation (big square).

2. Define the standard error of the estimate of the mean (again, in terms of our data and what it tells us).

   Answer: The SE of the estimate of the mean tells us how good our estimate of the mean (# of rocks per little square) is (AKA how much uncertainty there is in our estimate). It is calculated from the sample standard deviation divided by the square root of the number of sample units (little squares sampled).

3. How do sample standard deviation and standard error differ? (If you've explained that above, reiterate it here.)

   Answer: The sample standard deviation tells us something inherit about the population–the amount of variation in the number of dark rocks per little square. On the other hand, the standard error tells us how much uncertainty we have in our estimate of the number of rocks in each little square.

4. What would happen to the sample standard deviation if we sampled 20 sampling units (little squares) instead of only 8? Would it increase, decrease, or stay about the same? How about the standard error? Explain your rationale for both.

   Answer: The sample standard deviation estimate would stay about the same because it is an estimate of the amount of variation in the population, something that is a property of the population and won't change with increased sampling. The standard error, however, would decrease. That is because we now have more information about our population parameter that we are estimating (mean), so we have less uncertainty in the estimate (AKA smaller SE).

Set-up:

If you haven't already saved and shared your data from last week with your group:

(a) Get your (or your group's) simple random sampling rock data.
(b) Make sure it has been saved as a .csv file. If it hasn't, you will need to do so using "Save As".
(c) In Posit Cloud, using the "Upload" button in the lower right-hand panel, upload your .csv file.
(d) In the Files tab in the bottom right-hand corner, check the box next to the .csv file that you just uploaded, then click the "Rename" button (near "Upload").
(e) Rename your .csv file (something short, like "SRS_rocks.csv", if helpful)

Important Notes:

(a) I recommend referencing the PowerPoint slides from class! Remember, equations don't show on D2L, so you'll want to download the slides or use the PDF version.
(b) I also recommend taking a look back at Module 1 Assignment 3 to remind yourself how to work with data frames.
(c) Because each group will have different values for your number of rocks, the answer key won't directly match your values. I've included the example dataset that I used to create the answer key if you want to check your values.

Calculations

Each question is worth 1 point unless otherwise noted.

5. Use the `read.csv()` function from the tidyverse to read in your rocks file. Save this data in an object called `rocks`.

   Below, replace the file name with the file name of your rock data.

```
# this is for Ellen's code
rocks <- read.csv("../data_raw/example_data.csv")

# when you read in your file, the name of the file needs to be inside quotation marks
# it also needs to have the .csv at the end, as shown below
# rocks <- read.csv("your_file_name_here.csv")

# if your rocks data frame has more than 8 observations (check in the environment), run this line of co
rocks <- na.omit(rocks)
```

6. Create the object `N`, which is the number of units in your *sampling frame*. No functions needed; type the number and assign it to `N`.

Remember to print out the objects you create by typing the name of the object in the line below where you create it. This helps us out while grading quite a bit!

```
N <- 64
```

7. Create the object `n`, which is your *sample size,* the number of sample units that make up your sample.

   To do this, use the `length()` function to count the number of elements in the `sample_unit` column of the `rocks` data frame. Remember, we can use the `$` to reference a specific column in a data frame.

```
n <- length(rocks$sample_unit)
```

Estimating the Mean

One way we can calculate values is to call the column that we want and then save that value to an object.

For example, let's say we wanted to calculate the median of our dark stone abundance data. We could run the following line of code: `rocks_median <- median(rocks$abund_dark_stones)`

For the remainder of the coding questions below, remember to print out the answer by typing the name of the object again, below your line of code. In the above example, it would look like:

```
rocks_median <- median(rocks$abund_dark_stones)
rocks_median
```

8. Create an object called `rocks_mean` that contains the sample mean (AKA the mean number of dark stones per sample unit). Follow the code example above.

```
rocks_mean <- mean(rocks$abund_dark_stones)
```

9. Create objects called `rocks_var` and `rocks_sd` which contain the sample variance and sample standard deviation, respectively.

   Hint: use `var()` and `sd()` functions.

```
rocks_var <- var(rocks$abund_dark_stones)
rocks_sd <- sd(rocks$abund_dark_stones)
```

Calculating The Uncertainty in our Estimate of the Mean

R doesn't have a function for calculating the standard error (SE), so let's calculate it ourselves using the equation from class.

10. First, do you remember that finite population correction factor? We use that to modify SE when we know what N is. Let's create an object called `pop_correction` using that formula: (N-n)/N

    Hint: use the objects we've already created (`N`, `n`; don't enter any numbers directly)

```
pop_correction <- (N-n)/N
```

11. Why are we using the finite population correction factor in our formulas?

    Answer: We know the total number of sample units (64), so we need to adjust for that.

12. Using the objects you've already created and the formula (check the lecture slides from this module!), create an object called `rocks_var_ybar` which contains the variance of the *estimate* of the mean.

    Remember, this is different from the sample variance that you've already calculated, but that sample variance is included...

```
rocks_var_ybar <- pop_correction * (rocks_var/n)
```

13. Calculate the SE of the estimate of the mean by taking the square root of the `rocks_var_ybar` value. Call this object `rocks_SE`.

```
rocks_SE <- sqrt(rocks_var_ybar)
```

Estimating the Population (and Uncertainty of the Estimate)

We've calculated standard error (SE) for the estimate of the mean above. Now, let's calculate them for the estimate of the population abundance (tau hat)

14. First, let's calculate the total number of rocks in the population. Call this `rocks_total`. Use objects you've already created.

```
rocks_total <- N * rocks_mean
```

15. Now calculate the variance of the estimate of the population total, `rocks_var_total`.

```
rocks_var_total <- N * (N-n) * (rocks_var/n)
# alternative equation is N^2 * rocks_var_ybar
```

16. Calculate the SE of the estimate of the population total, `rocks_SE_total`.

```
rocks_SE_total <- sqrt(rocks_var_total)
```

Bringing It All Together

17. Write a short paragraph (3-5 sentences) putting the values that we've just calculated into context for our dark rocks sampling adventure. Be sure to answer the following questions. (3 points)

- According to your calculations, approximately how many dark rocks were there in your big square? How much uncertainty is in your estimate?
- What is your estimate of density of dark rocks (average # dark rocks per little square)? What is your level of uncertainty is your estimate in the density? How much variation is in the number of dark rocks per little square throughout your large square?

Answer:

- number of dark rocks in big square = `rocks_total`
- confidence (uncertainty) is estimate of number of dark rocks total = `rocks_SE_total` or `rocks_var_total`
- density estimate = `rocks_mean`
- confidence (uncertainty) in density estimate = `rocks_var_ybar` or `rocks_SE`
- variation in density = `rocks_var` or `rocks_sd`