Reproducible Research with R

In-course exercise

Brooke Anderson

1/27/2021

Overview

Overview

The in-class exercise will have two components:

1. Some basic testing and exploration of creating and editing documents with RMarkdown (10–15 minutes)
2. Start building a protocol for pre-processing data from plating to measure the concentration of viable bacteria in samples (~ 25 minutes)

The second you will continue to work on with your group outside of class. You will submit the result as your assignment for this week (due next week, I believe).

Groups

The groups for today are:

- **Group 1:** Sere, Tyler, Tyler, Daniel, Pablo
- **Group 2:** Ikaia, Joey, Hector, Kaz, Naly
- **Group 3:** Camron, Emily, Sean, Carlos, Roland
- **Group 4:** Casey, Kristen, Alissa, Gaby, Lexi

Basic exploration

Working together in your group

- Plan to spend about 10–15 minutes on this part—it's really just to get a feel for making and editing RMarkdown documents
- For this part, let those new to RMarkdown take the lead, and have one share his or her screen.
- Team members who have used RMarkdown please guide those on your team who are new and are trying this up while sharing their screen.

Basic exploration of RMarkdown

For this first part of the exercise, do the following:

1. Open RStudio. Create a new RMarkdown document that will render to Word.
2. Once the document opens, save the file, and then click on the "Knit" button to render the document to Word. You should get a new output file in Word that's the result of rendering the RMarkdown.
3. Look at the RMarkdown document and the rendered Word document. See if you can connect each part of the Word document to the part in the RMarkdown document that created it.

Basic exploration of RMarkdown

4. Change some things in the RMarkdown document. For example, delete some paragraphs or sections. Change some of the code in the code chunks. Use the RMarkdown cheatsheets—add some text with special formatting and try changing some code chunk options. Re-render the document (with the "knit" button).
5. Change something "by-hand" in the Word document. Re-render the RMarkdown document. Did it keep your change, or was it overwritten? Discuss how this relates to the idea of "read only".
6. Try using the green arrow in one of the code chunks to run the code just in that section. Try using the small arrow next to the "Knit" button to render the document to HTML, instead of Word.

Pre-processing protocol for plating data

Pre-processing protocol for plating data

For the assignment, we are asking you to develop an RMarkdown document that includes the code an description for how you **pre-processed** plating data, to get from CFUs at different dilutions for a sample to an estimate of the bacteria concentration in the sample.

- You will complete this assignment in the groups you are in today.
- Each person should submit their own copy of the group's final document through Canvas (but we expect the final documents to be identical across group members). You should submit both the RMarkdown file (".Rmd") and the output document (".docx", ".html", or ".pdf").

Colony-forming units (CFUs)

You may be testing out some drugs against an infectious bacteria and want to know how successful different drugs are in limiting bacterial load. You run an experiment and have samples from animals treated with different drugs or under control.

You want to know: How much viable (i.e., replicating) bacteria are in each of your samples?

You can find out by **plating** the sample at different **dilutions** and counting the **colony-forming units (CFUs)** that are cultured on each plate.

Plating

You put a sample on a plate with a medium they can grow on and then give them time to grow.

The idea is that individual bacteria from the original sample end up randomly around the surface of the plate, and any that are viable (able to reproduce) will form a new colony that, after a while, you'll be able to see.

Dilution for counting

To count the number of colonies, you need **a "just right" dilution** (likely won't know what this is until after plating) to have a **countable plate**.

If you have **too high** of a dilution (i.e., one with very few viable bacteria), randomness will play a big role in the CFU count, and you'll estimate the original with more variability.

If you have **too low** of a dilution (i.e., one with lots of viable bacteria), it will be difficult to identify separate colonies, and they may complete for resources. (The pattern you see when the dilution is too low (i.e., too concentrated with bacteria) is called a *lawn*—colonies merge together into blobs.)

Dilution for counting

To translate from diluted concentration to original concentration, you need to do a back calculation.

You have found:

- The "just right" dilution
- The number of colonies that grew in that dilution

You need to use these numbers to estimate the concentration of viable bacteria in the original sample.

Example data

You will be working with example data from a package a student at CSU (Amy Fox) is developing called `bactcountr` (https://github.com/aef1004/bactcountr).

It provides data from plating samples from an experiment. The goal is to estimate the concentration of *Mycobacterium tuberculosis* in two organs (spleen and lung) of mice.

Example data

It is available to download at https://github.com/aef1004/bactcountr/tree/master/data. Download this file and save it to the same directory as your RMarkdown for this assignment.

Because the data are in a binary R file, you can load it in your R session just by using the `load` function (see next slide for example code).

Example data

Here are the first few rows:

```
load("data/CFU_raw_formatted.rda")
head(CFU_raw_formatted, 3)
```

```
##   group replicate  organ dilution CFUs
## 1     2         A Spleen        0   26
## 2     2         C Spleen        0    0
## 3     3         A Spleen        0    0
```

This dataset has already been converted to a "tidy" format, to make it easier to work with in R. Each row gives the counted CFUs for one plate.

Example data

It includes columns for:

- `group`: The experimental group that the animal was assigned to
- `replicate`: The animal replicate. The rows with the same values for *all* of group, replicate, and organ come from the same original sample.
- `organ`: Which organ of the animal the sample came from
- `dilution`: The level of dilution for the plated sample (with a dilution factor of 5, i.e., 5-fold dilutions at each step)
- `CFUs`: The colony-forming units (CFUs) counted on that plate.

Missing values in the `CFUs` for a certain dilution mean that the CFUs were too numerous to count at that dilution.

Example data

Since samples are plated at multiple dilutions, there are several rows per original sample:

```
library(tidyverse)
CFU_raw_formatted %>%
  filter(group == 2 &
           replicate == "A" &
           organ == "Spleen")
```

```
## # A tibble: 4 x 5
##   group replicate organ  dilution  CFUs
##   <dbl> <chr>     <chr>     <dbl> <dbl>
## 1     2 A         Spleen        0    26
## 2     2 A         Spleen        1    10
## 3     2 A         Spleen        2     0
## 4     2 A         Spleen        3     0
```

Pre-processing tasks

You'll need to do the following general pre-processing tasks with this data, describing and showing each in the RMarkdown document for this week's assignment.

1. Load the data into your R session
2. Explore the data, to show its dimensions and characteristics and identify any potential quality control issues in the data
3. For each sample (measurement for one organ in one experimental animal), identify a good dilution to use for estimating the CFUs in the original sample
4. Use the CFUs counted for this "just right" dilution to estimate the bacterial concentration in the original sample
5. An explanation of the final data's format (e.g., what columns does it have, what's in each column)

Pre-processing tasks

You may need to do some work to figure out a good way to do steps 3 and 4.

Draw on any background knowledge you have.

Also, feel free to use any resource you can think of to figure out a reasonable way to do these steps.

Regardless of what you do and how you figure out how to do it, be sure to document your choices clearly and thoroughly in the RMarkdown document.

Elements in the final document

Each document should include:

- A meaningful title, the date the assignment is due, and the names of all team members as authors (this should all be done in the YAML).
- Section headers for each section (there should be 5 sections, as listed above in the five pre-processing tasks)
- At least two references to journal articles or books to justify choices for how you've pre-processed the data.
- At least one formatted mathematical equation / expression to explain your calculations
- Clear and thorough explanations for how you determined a good dilution for each sample and how you calculated the concentration in the original sample using that dilution.
- Minimal typos—use RMarkdown's spellcheck button!

Pre-processing tasks

You may assume that:

- The whole organ was used as the sample, and we are trying to estimate bacterial concentration in this whole organ
- In each case, the volume plated is 100 microliters
- The volume used to resuspend the CFU solution was 0.5 milliliters in each case
- The dilution factor is 5
- You may calculate either whole CFUs or log CFUs, just make sure you're clear about which you present in your final data

If you need to make other assumptions, that's no problem—but make sure that you state any you make in the RMarkdown document so this is documented.

Pre-processing tasks

Some `tidyverse` functions I found helpful:

- `mutate`
- `abs` (takes an absolute value)
- `group_by` (you can group by more than one variable, so you can group all platings from the same original sample together)
- `arrange` (you can also arrange by more than one thing—it will only use the later variables for arranging if there are ties based on earlier ones),
- `slice`

Resources

Resources on BibTex in RMarkdown:

- https://bookdown.org/yihui/rmarkdown-cookbook/bibliography.html
- https://rmd4sci.njtierney.com/citing-articles-bibliography-styles.html
- https://rmarkdown.rstudio.com/authoring_bibliographies_and_citations.html

Resources on math in RMarkdown:

- https://rmd4sci.njtierney.com/math.html
- https://rpruim.github.io/s341/S19/from-class/MathinRmd.html
- https://www.stat.cmu.edu/~cshalizi/rmarkdown/#math-in-r-markdown