Protocol: Estimating bacterial loads from plating samples at different dilutions

3/1/2021

Overview

This protocol provides details on how you can estimate the bacterial load in samples based on data from plating those samples at different dilutions. The question being answered through this protocol of data analysis is: 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. 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.

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). To translate from diluted concentration to original concentration, you can then do a back-calculation, incorporating both the number of colonies counted at that dilution and how dilute the sample was.

Laboratory description

These data result from plating each sample at several different dilutions and counting the colony-forming units (CFUs) that are cultured on each plate. Each sample was on a plate with a medium they can grow on and then given [time] to grow. On each plate, the individual bacteria from the original sample should have ended up spread randomly around the surface of the plate, and any that are viable (able to reproduce) formed a colony that, by the end of the procedure, can be counted.

[Image of plating]

Each sample was taken from a single mouse from the experiment. These mice came from different experimental groups, with several replicates (mice) per experimental group. The experimental groups in this case were:

[table of experimental groups]

Data description

Data are collected for this process into a spreadsheet that is set up in a "tidy" format used by our laboratory to collect this type of data. Here are the first few rows of the data:

```
## # A tibble: 6 x 4
##   group replicate dilution  CFUs
##   <dbl> <chr>        <dbl> <dbl>
## 1     2 2-A              0    26
## 2     2 2-C              0     0
## 3     3 3-A              0     0
```

```
## 4     3 3-C              0     0
## 5     4 4-A              0     0
## 6     4 4-B              0     0
```

Each row represents the number of bacterial colonies counted after plating a certain sample at a certain dilution. Columns are included with values for the experimental group of the sample (`group`), the specific ID of the sample within that experimental group (`replicate`, e.g., `2-A` is mouse A in experimental group 2), the dilution level for that plating (`dilution`), and the number of bacterial colonies counted in that sample (`CFUs`).

Reading data into R

The data are stored in a comma-separated plain text file called "cfu_data.csv". They can be read into R using the following code:

```
library(tidyverse)
cfu_data <- read_csv("cfu_data.csv")
head(cfu_data)
```

```
## # A tibble: 6 x 4
##   group replicate dilution  CFUs
##   <dbl> <chr>        <dbl> <dbl>
## 1     2 2-A              0    26
## 2     2 2-C              0     0
## 3     3 3-A              0     0
## 4     3 3-C              0     0
## 5     4 4-A              0     0
## 6     4 4-B              0     0
```

You will need to be sure that the data file ("cfu_data.csv" in this case) is in the working directory of your R session. You can check the current working directory in R with the call `getwd()`, and you can list all files in the current working directory with the call `list.files()`. These function calls can be useful to check to make sure that this file is in your working directory and, if not, to identify which directory on your computer file system you should move it to.

Once you run this command, the data will be available in your R session in the object `cfu_data`. You can see the first few rows by running:

```
head(cfu_data)
```

```
## # A tibble: 6 x 4
##   group replicate dilution  CFUs
##   <dbl> <chr>        <dbl> <dbl>
## 1     2 2-A              0    26
## 2     2 2-C              0     0
## 3     3 3-A              0     0
## 4     3 3-C              0     0
## 5     4 4-A              0     0
## 6     4 4-B              0     0
```

You can get a summary of the data by running:

```
summary(cfu_data)
```

```
##      group         replicate            dilution          CFUs
##  Min.   : 2.000   Length:89          Min.   :0.000   Min.   : 0.000
##  1st Qu.: 4.000   Class :character   1st Qu.:1.000   1st Qu.: 0.000
##  Median : 5.000   Mode  :character   Median :2.000   Median : 0.000
```

```
##  Mean   : 5.933                    Mean   :1.539   Mean   : 1.944
##  3rd Qu.: 8.000                    3rd Qu.:3.000   3rd Qu.: 0.000
##  Max.   :10.000                    Max.   :3.000   Max.   :52.000
```

Exploring the data and quality checks

It is helpful to explore the data once you read it in, to check for quality control issues and other characteristics of the data from a particular experiment. This section provides code examples for this quality control.

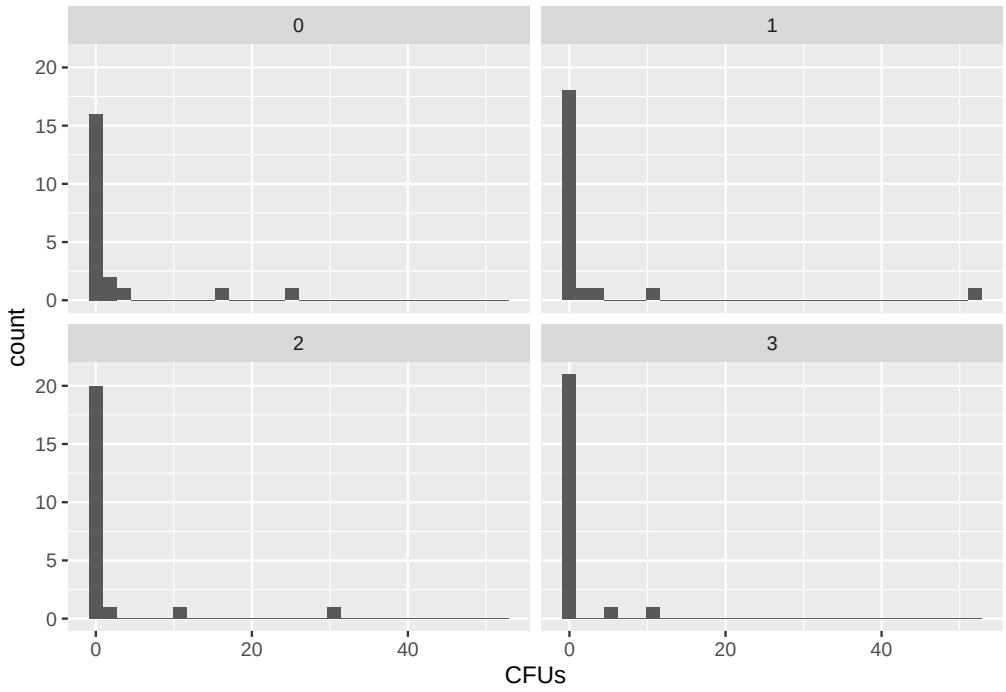
First, it can be helpful to ensure that the data includes the number of experimental groups, and the number of replicates within each group, that you are expecting. First, we can determine that there are eight experimental groups, with between eight and twelve samples (replicates) in each group:

```
cfu_data %>%
  group_by(group) %>%
  count()
```

```
## # A tibble: 8 x 2
## # Groups:   group [8]
##   group     n
##   <dbl> <int>
## 1     2    11
## 2     3    10
## 3     4    12
## 4     5    12
## 5     7    12
## 6     8    12
## 7     9    12
## 8    10     8
```

It is also helpful to see the distribution of CFUs at each dilution level, which can be plotted with the following code:

```
cfu_data %>%
  ggplot(aes(x = CFUs)) +
  geom_histogram() +
  facet_wrap(~ dilution)
```



In this case, most bacterial counts are zero at all dilution levels. However, there are more non-zero CFUs at dilution level 0 and more as you move to higher dilution levels. The highest CFU count at any dilution level is a little over 50, for dilution level 1. Most non-zero CFU counts are at or below 30, regardless of dilution level.

Identifying a good dilution for each sample

To count the number of colonies, you need a "good" dilution to have a **countable plate**. You won't be able to pick a "good" dilution for each sample until you have plated several dilutions and see approximately how many colonies form on each. This is why a plating experiment will include plates for each sample at several dilution levels of the sample.

It is important to use a "good" dilution to estimate the bacterial load in the original sample, because there are challenges and pitfalls to trying to count CFUs if the dilution is too high or too low.

If you try to count for a plate that has too high of a dilution (i.e., the sample has been diluted a lot, and so has very few viable bacteria), randomness will play a big role in the CFU count, and you'll estimate bacterial load in the original sample with more variability than desired. If you have too low of a dilution (i.e., one that has not been diluted much and has 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).

Typically, there will be some standard range of CFUs that are indicative of a plate having a "good" dilution for counting. For example, a laboratory group may specify that plates with between [x] and [x] CFUs are ideal for counting CFUs, and so use the dilution for each sample that provides a count in this range when estimating bacterial load in the original sample. The exact range used to identify plates with "good" dilutions may range from laboratory group to laboratory group. It may also depend on the type of bacteria being plated.

In this protocol, we are describing how to analyze data from experiments with *Mycobacterium tuberculosis*. This is a very slow-growing bacteria, and so we typically identify "good" dilutions using a lower range of CFUs than other laboratories might when working with faster-growing bacteria. The trade-off, of course, is that there will likely be more variability in our bacterial load estimates, since they tend to be based on plates with lower average counts. However, we balance this limitation with the practical limitation of how long it

takes for this type of bacteria to grow, and so need to settle for using lower CFUs in some cases than might be ideal if working with faster-growing bacteria.

Estimating bacterial load for each sample

Outputing final estimates