Species richness analysis

Identifying species in microbial communities

To describe the species found in a microbial community, we isolate the DNA of a sample and then amplify and sequence a suitable barcoding marker, typically the 16S rRNA gene. We then group all similar sequences into so-called **amplicon sequence variants (ASV)**. For our purposes here, we assume that *each ASV corresponds to a bacterial species*. By counting the number of each ASV in each sample, we get an ASV counts table (Table 1).

Table 1: ASV counts of samples collected at different depths of a lake.

```
   name depth_10_m depth_20_m
1 ASV_1         28         14
2 ASV_2          5         12
3 ASV_3          0          3
4 ASV_4         14         41
5 ASV_5         18         28
```

Species richness

A simple way of characterizing a microbial community is to count how many species/ASVs we have in a sample. This metric is called **species richness**. For example, to determine the species richness of the water sample at 10 m, we can select just that column, filter the rows for ASVs with one or more counts, and then count the number of rows.

```
species_richness_10_m <- asv_counts |>  # Use the ASV counts table.
  select(depth_10_m) |>                 # Select the sample of interest.
  nrow()                                # Count the number of rows.
```

The water sample at 10 m has a species richness of 5.