Case Study 1

BIOE 498/598 PJ

Due: 2/24/2021 before 5pm Central

Problem Statement

Some bacteria use *natural competence* to take up DNA from their environment. In streptococci, natural competence is controlled by a gene called *comR*. Recently it was discovered that the species *Streptococcus sobrinus* has two copies of the *comR* gene — *comR1* and *comR2*. Your goal is to investigate which of the two genes (if any) regulates competence.

Transformation assays can be used to assess natural competence. A plasmid containing an antibiotic resistance gene is added to a culture. The bacteria are given time to take up the DNA before they are plated on agar containing antibiotics. The efficiency of transformation can be quantified by counting the number of colonies formed in the presence of the antibiotic.

Your dataset includes transformation results for combinations of three different genotypes for *comR1* and *comR2*. The wildtype background is the unmodified bacterium. In knockout strains a gene has been deleted from the genome. It is customary in microbiology to *complement* the knockout strains by adding another copy of the gene, in this case on a plasmid. The bacteria maintain several copies of the plasmid, so complemented strains contain multiple copies of either *comR1* or *comR2*, but the exact copy number is unknown.

Your goal is to answer two questions. First, what are the roles of *comR1* and *comR2* in regulating natural competence? Second, do the complemented strains behave like the wildtype strains, or do the additional copies of the gene change the transformation efficiency?

Loading the data

Run the following code to load a dataframe with the transformation results.

```r
data <- read.csv("comR12_tx_data.csv")
data$comR1 <- gdata::reorder.factor(data$comR1, new.order=c("wt", "ko", "oe"));
data$comR2 <- gdata::reorder.factor(data$comR2, new.order=c("wt", "ko", "oe"));
```

The second and third lines reorder the factors so the wildtype (`wt`) is the base factor level. Without reordering, R would sort the levels alphabetically and choose the first level as the base.

```r
head(data)
```

```
##   comR1 comR2 efficiency block
## 1    wt    wt  22.500000   F21
## 2    wt    wt  27.049180   F21
## 3    wt    wt   8.522727   F21
## 4    wt    wt  51.282051   F22
## 5    wt    wt  18.154312   F22
## 6    wt    wt  32.106782   F22
```

The dataframe has four columns:

- `efficiency` is the response variable: the number of colonies seen when a culture of 10^6 cells was transformed and plated under antibiotic selection.
- `comR1` and `comR2` are the genotypes of the strains, either wildtype (`wt`), knockout/deletion (`ko`), or a complemented strain with additonal copies of the gene (`oe` for over-expressed).
- `block` is a blocking factor indicating the date of the experiment.

Questions

1. Build a linear model that predicts transformation efficiency from genotype, including any potential interactions between *comR1* and *comR2*.
2. Determine if a transformation of the response variable would improve your model's predictions. If so, perform the transformation.
3. Make a *predicted vs. actual* plot for your data by plotting the model's predictions for every run in the dataset against the measured transformation efficiency. If you transformed your response variable, make a separate plot for the model with this transformation. (Useful functions: `plot` and `predict`.)
4. Does *comR1* affect transformation efficiency? How about *comR2*? Do these genes interact?
5. Our concern is that our complementation stratey adds too many copies of the gene. Does the `oe` level differ from the `wt` level for either *comR1* and *comR2*?

Format

You should answer the questions by creating a set of slides. Imagine you are presenting the results at an internal group meeting of microbiologists. Include any summary information about the dataset as well as your conclusions. You may include supplementary slides with your code and model output, but the main slides should present the analysis plan and results in a format accessable to scientists without DOE training.