

This assignment will provide experience with repeated measure analysis of variance (rm ANOVA).

Use the “log CFU data” dataset for this assignment. This is data on 80 observations in 20 animals, 4 observations per animal, one at each of 4 times.

In this study, animals were infected with Tuberculosis (TB) and assigned to an old “proven” antibiotic treatment ($tx=0$) or new treatment ($tx=1$). Animals were measured at baseline (month 0) before treatment was started, and subsequently at months 1, 2 and 3 after treatment began. At each month, a small sample of fluid was removed from their lungs and TB was cultured for 24 hours. After 24 hours, the numbers of colony forming units (CFUs) were measured. The CFU can be a very large number. The log base 10 CFU ($\log_{10}CFU$) are given in the dataset and is the outcome (Y) variable.

You are to compare the mean log CFU values across time between the two groups. Report whether the mean log CFU values across time are significantly different and in what direction. You may wish to make a plot of mean log CFU versus months by group. Investigate whether the relation between month and mean log CFU is reasonably approximated by some simple function of time, such as a linear or quadratic function, and whether the change over time is parallel between the two treatment groups. (Technical question, based on what we have learned, can one compute an exact test and p value for comparing two models that are NOT nested? Is a linear regression model for time nested in an ANOVA model for time?).

Investigate whether a correction should be made for any lack of comparability of baseline (month 0) log CFU values. If such a correction is needed, you should carry this out.

Based on your findings for mean log CFU, you are to also report on whether mean CFU on the **original scale** is a simple (linear, quadratic, ?) function of month for each group. For example, if mean log CFU is a linear function of time, is mean CFU also a linear function of time?

You should include graphs and descriptive statistics tables that include summary information (sample size, means, standard errors of the means, etc.) across time for each treatment group. Indicate what model, if any, these calculations are based on. You should also comment on whether the design is balanced and whether the model based means are the same as the means that would be obtained in the absence of a model (the purely descriptive means).