Write an R script to solve the problems prefixed with an R. In Word (or any other word processor), answer all the problems prefixed with a W, without using R. Then convert the word processed file (not the R script) into pdf format. Submit both files, named according to the convention given in Homework 0, attached to a single email.

Write an R script for the following.

Information on this data set from its source, https://dasl.datadescription.com: “A student investigated just how effective washing with soap is in eliminating bacteria. To do this she tested four different methods—washing with water only, washing with regular soap, washing with antibacterial soap (ABS), and spraying hands with antibacterial spray (AS) (containing 65% ethanol as an active ingredient). Her experiment consisted of one experimental factor, the washing Method, at four levels. She suspected that the number of bacteria on her hands before washing might vary considerably from day to day. To help even out the effects of those changes, she generated random numbers to determine the order of the four treatments. Each morning, she washed her hands according to the treatment randomly chosen. Then she placed her right hand on a sterile media plate designed to encourage bacteria growth. She incubated each plate for 2 days at 36°C, after which she counted the bacteria colonies. She replicated this procedure 8 times for each of the four treatments.”

We are interested in the question: On average, how do the bacteria counts from the three other methods of hand-washing compare to those of the control (water only, no soap)?

To address this question, we define the random variable $B$ that is the bacteria count as measured by the process described above, and the variable that is the method $M$ on hand-washing used during the process.

The count column in the data frame contains our observations of the bacteria counts, and the method column is a categorical variable with four levels, ordered as: Control (Water), Soap (Antibacterial), Soap (Regular), Spray (Alcohol).

It would be good to think about what an individual is here, and what population our statistical inferences extend to.

R1. Produce simultaneous density plots of the counts produced by each method.

R2. Fit a model with model formula $B \sim M$, and produce a model summary of the fitted model.

R3. Produce a simultaneous density plot of the standardized residuals of the fitted model. (The axes don’t need any particular labels, but there should be a legend.)

R4. Produce a normal quantile plot of the residuals of the fitted model.

Answer the following, but don’t include any R code in your write-up.

W1. With $B$ and $M$ defined as above, how should we denote the true model equation with model formula $B \sim M$ to address our question of interest?

W2. We check the model fitting assumptions, so we fit a model with model formula $B \sim M$. The model summary for this model (which you should produce yourself in the R section as well) is
What is the model equation for this fitted model?

W3. The sampling variability assumptions look fine (about as nice as one sees for such models!), so we proceed to conduct statistical inference. Report and interpret each of the \( t \) tests conducted as part of the model summary. In answering this question, as usual state the null hypothesis of each test (but you don’t need to include the alternative hypothesis), and include the value of the test statistic, the degrees of freedom, the \( p \)-value, and what this did or did not constitute statistically significant evidence against. Your answer should be statistical in its language, using terms such as \( \mu[B|M = \text{Soap (Regular)}] \), for example.