

Assignment: Your task is to analyze the associated data set (**Plant_stress.tsv**) as if you generated it and are presenting the results for publication. You should use R for your work, and produce the appropriate analyses and figures to test distributional assumptions and to support your conclusions. You should encapsulate these findings in a results section that is modeled after one you would find in a journal such as Science or Nature, or PLoS Biology. You should not present every intermediate figure and analysis! Remember, the goal is to concisely and accurately present your results in complete figures, tables and statistics, each of which is correctly labeled and annotated. You may have additional figures that you can present as supplementary material, but be sure that they are referenced as such. In addition, include your final R script (.Rmd file) as supplementary material.

Due: Submit your work **via Canvas** by the end of the day (midnight) on **Sunday November 26th**. You can work with other members of class, but I expect each of you to construct and run all of the scripts yourself, and to write the results section on your own.

The Study:

Pretend that you performed a study to determine whether or not the levels of three different proteins measured by mass spectrometry (**ERD10, NCED9 and SRK2D**) in the root tissues of *Arabidopsis* plants are affected by two different factors (**Genotype and Water**). Values for the 3 proteins correspond to signal intensity units. The **Water** factor is a treatment that you applied at two levels: **control**, which is the normal amount of daily water or a small amount of water 1/10 that of normal (**stress**). In addition, you performed the study on two different genotypes of plant (**Genotype**), those with a loss-of-function mutation in a gene encoding a heat shock chaperone protein (**mutant**) and those without the mutation (**wildtype**).

Here are some **things to keep in mind - but not necessarily include in your results** - to help you prepare your results section:

1. What type of experimental design is this, what class of ANOVA should you run, and what are all of the null and alternative hypotheses relevant to your question for each of the proteins?
2. What are the assumptions about your analysis, and how do you test these assumptions?
3. How should the factors be treated in your analysis models (think fixed vs. random effects)?
4. What statistics should you present for each hypothesis test?
5. How should you interpret your results, and what type(s) of figures would help with your interpretation of the different effects in the model?
6. Do your interpretations of the effects of the two factors differ for the three different proteins?