

Extra Practice for Exam2

1. An investigator is interested in comparing the expression of a certain gene for plants grown under 2 different Conditions (Trt and Ctrl). 4 seedlings were randomly assigned to the Ctrl condition and 4 seedlings were randomly assigned to the Trt condition. So, there are 4 Plants per Condition. At the end of the study period, an RNA sample was obtained from each plant and split into triplicates (labeled Rep in the data). Each sample was analyzed using RT PCR and gene expression (Y) was measured. Due to the triplicates, there were 3 observations for each Plant for a total of 24 observations (2 x 4 x 3). **Note:** We will consider Condition to be fixed and Plant to be random. The data is available from Canvas as “**Plants.csv**”
 - A. Are Plant and Condition crossed or nested? If nested, be sure to indicate the “direction” of the nesting.
 - B. Considering your answer to A, fit an appropriate model. Include the variance parameter estimates and Type 3 ANOVA table in your assignment.
 - C. Using the model from part B, can we conclude there is a difference between the mean responses for the two Conditions? Provide an estimate of the difference and a p-value. Hint: Use emmeans.
 - D. We will rerun the analysis using a different approach. We will start by averaging over the triplicates and run a two-sample t-test (assuming equal variance) with n=4 observations per Condition. To do this, use code similar to below:

```
SumStats <- aggregate(Y ~ Condition + Plant, data =  
Plants, FUN = mean)  
t.test(Y ~ Condition, var.equal = TRUE, data = SumStats)
```

In your assignment, provide an estimate of the difference and a p-value. Note: Using option var.equal = TRUE, returns the two sample t-test assuming equal variance.
 - E. Compare your result from part D to the test from part C. Are the results the same?**For parts F and G**, a colleague suggests that you “just do a two-sample t-test with n=12 observations per condition”.
 - F. Would this analysis be appropriate? Justify your response.
 - G. Would this analysis give the same results as part C? Hint: Try it and compare the results!
2. A **split-plot** experiment was conducted on sorghum with two treatment factors: plant Density and Hybrid. A total of 4 blocks were used in the study. Within each block, the four levels of plant Density (10, 15, 25 and 40 plants per meter of row) were randomly assigned to whole plots. Then within each whole plot, the three Hybrids (A, B, C) were randomly assigned to subplots. The response (Y) is the weight of the seed per plant in grams. The data are given in the file “**SeedWeight.csv**” on Canvas.
 - A. Create an interaction plot (for Density and Hybrid) and include it in your assignment. For consistency, put Density on the horizontal axis.
 - B. Fit an appropriate model using lmer(). Include the variance component estimates and Type 3 ANOVA table in your assignment.

- C. You should find that the Density*Hybrid interaction is not significant. Use `emmeans()` to run pairwise comparisons of Density levels (averaging over Hybrids) and comparisons of Hybrids (averaging over Density levels).
- D. On slide 61 (Random2 notes), it says “the accuracy on factor B (sub-plot factor) is better than the accuracy of comparisons on factor A (whole-plot factor)”. Looking at your results from part B, what evidence do we see to support this statement (for this dataset).
- E. Use `plot()` to generate the plot of residuals versus fitted values. You do not have to include this plot in your assignment, just comment on what you see.