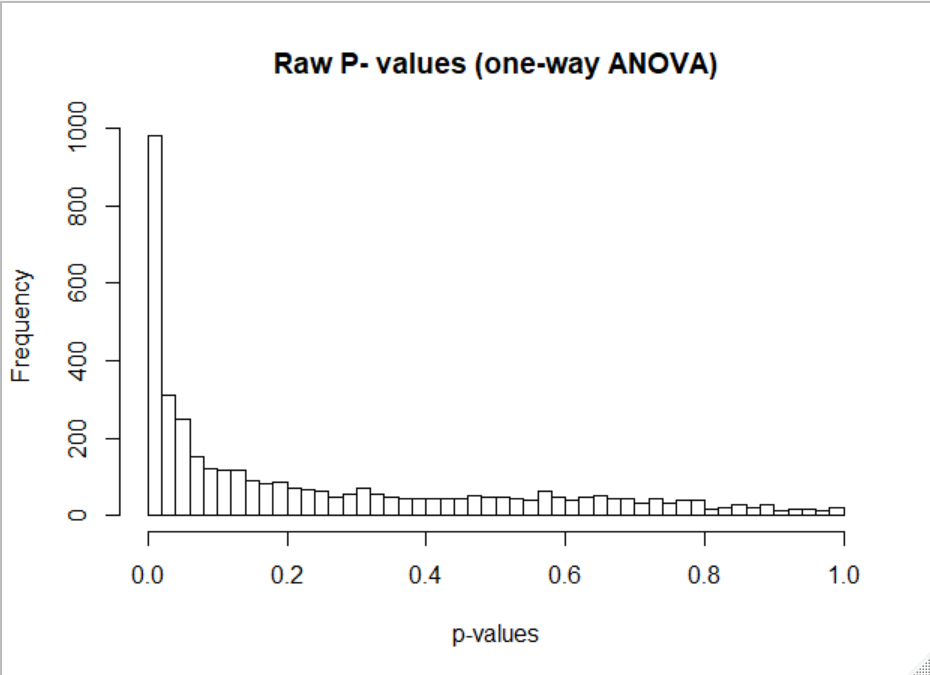
(The first 3 columns are “day 2”, the next 3 columns are “week 12” and the last 5 are “week 18”).

1. For each row in the spreadsheet, perform a one-way ANOVA with categories “day 2”, “week 12” and “week 18”. Plot out the histogram of all p-values. How many genes are significant at a BH FDR-corrected 0.05 threshold. (see mini-lecture 16B).

**I executed the command at listing 1 to calculate the number of BH FDR-corrected significant genes (612) at 0.05 threshold and to generate the histogram at figure 1. The source file is located at URL** <https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab10/onewayanova.R> .

  
**Listing 1: R command for one-way ANOVA**



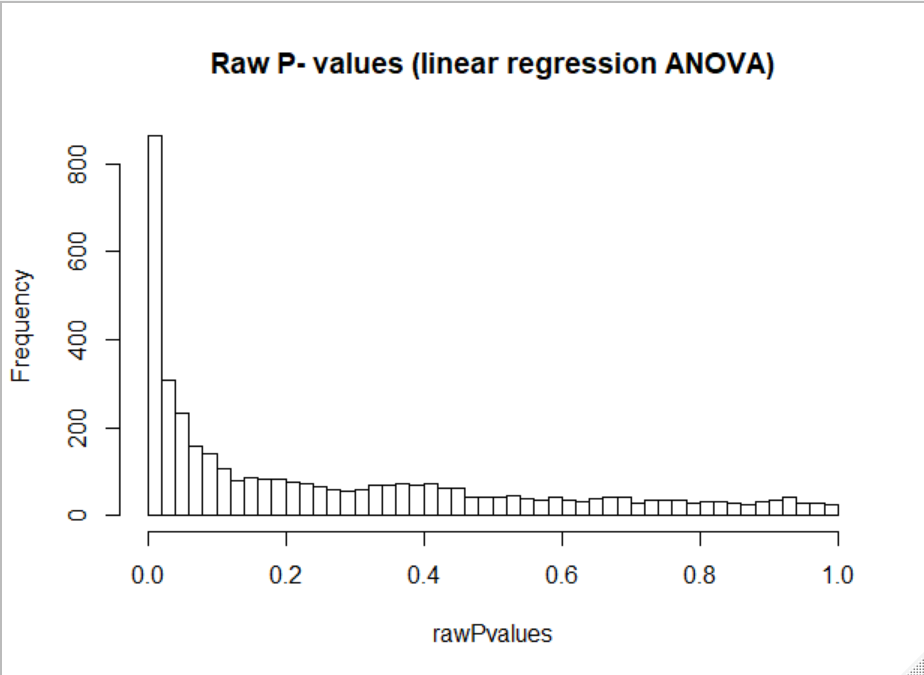
**Figure 1: Raw p-values for one-way ANOVA**

1. Next make an ANOVA as a linear regression as a function of time (so 2 days, 86 days and 128 days). Plot out the histogram of all p-values. How many genes are significant at a BH FDR-corrected 0.05 threshold. (see lecture 15)

**I executed the command at listing 2 to calculate the number of BH FDR-corrected significant genes (448) at 0.05 threshold and to generate the histogram at figure 2. The source file is located at URL** [**https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab10/linearanova.R**](https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab10/linearanova.R) **.**



**Listing 2: R command for ANOVA as a linear regression of a function of time**



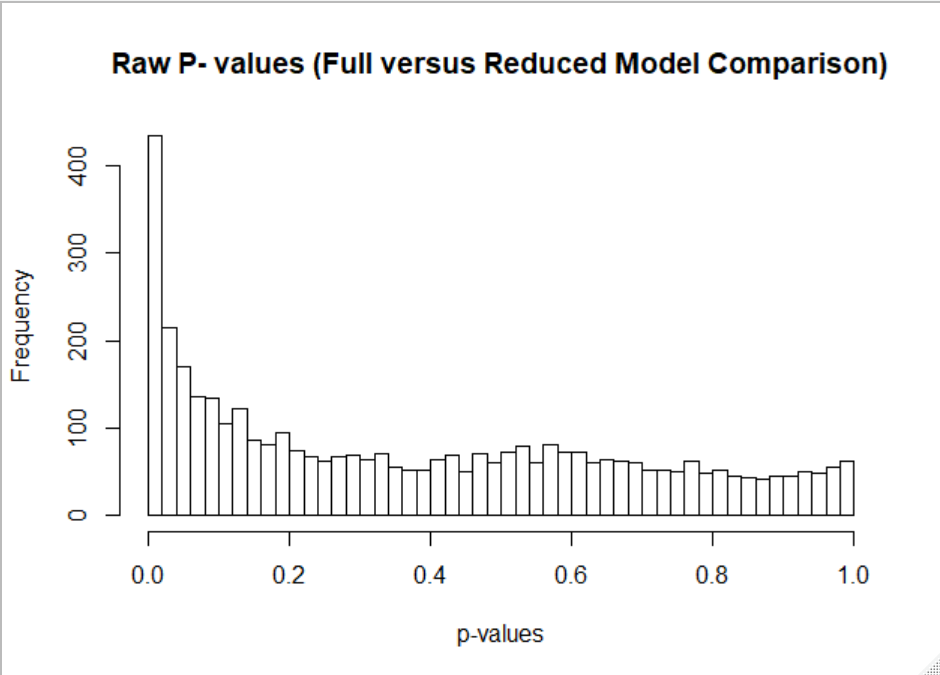
**Figure 2: Raw p-values for ANOVA as a linear regression of a function of time**

1. Finally, for each row in the spreadsheet perform an ANVOA comparing the three-parameter model from (A) and the two parameter model from (B). (see mini-lecture 16C). Plot out the histogram of all p-values. For how many genes is there a significant difference between these two models at a BH FDR-corrected threshold.

**I executed the command at listing 2 to calculate the number of BH FDR-corrected significant genes (51) at 0.05 threshold and to generate the histogram at figure 2. The source file is located at URL** [**https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab10/mixedanova.R**](https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab10/mixedanova.R) **.**



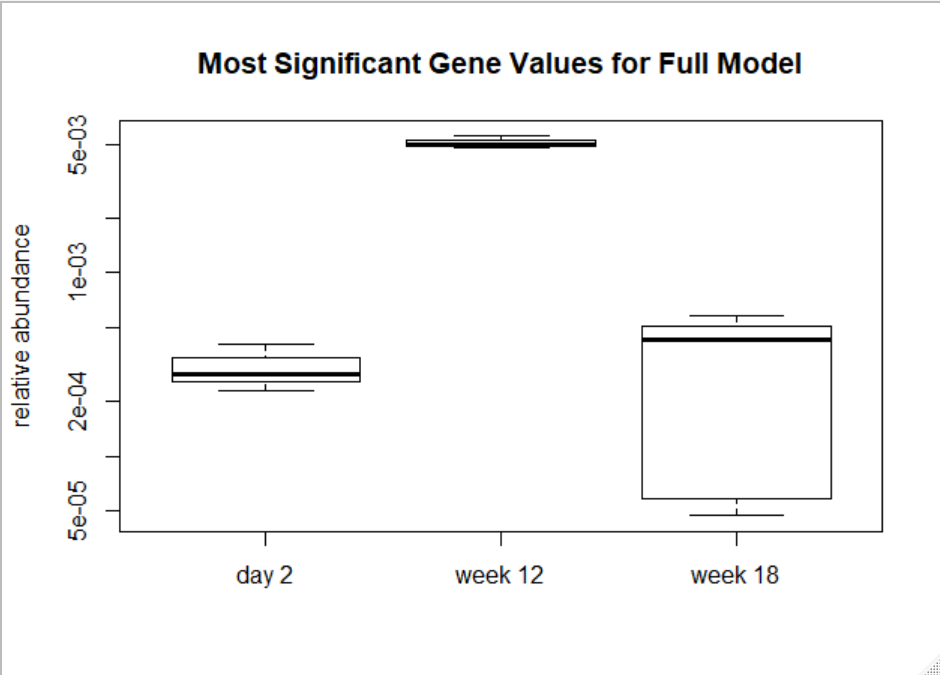
**Listing 3: R command for ANOVA full versus reduced model comparisons**



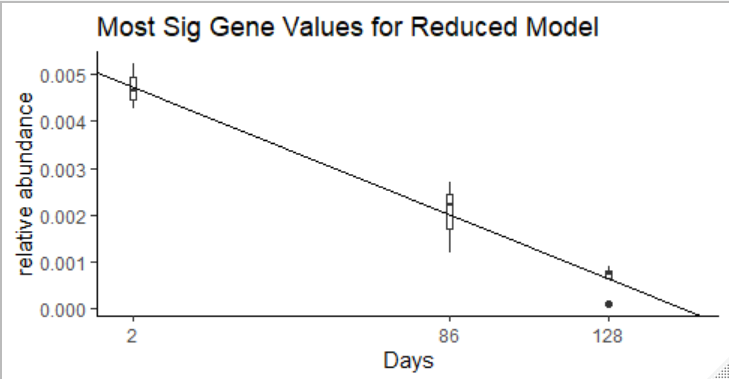
**Figure 3: Raw p-values for ANOVA full versus reduced model comparisons**

1. Make three graphs showing the relative abundance of the most significant gene under each of the three ANOVA models. For (A) and (C), the x-axis will the category (day 3, week 12 and week 18) and the y-axis will be the relative abundance. Be sure to properly label and title all graphs and axes. For (B) the x-axis will be time (in days) and the y-axis will be the relative abundance. For the graph of the top hit from (B), include the regression line for the plot from (B).

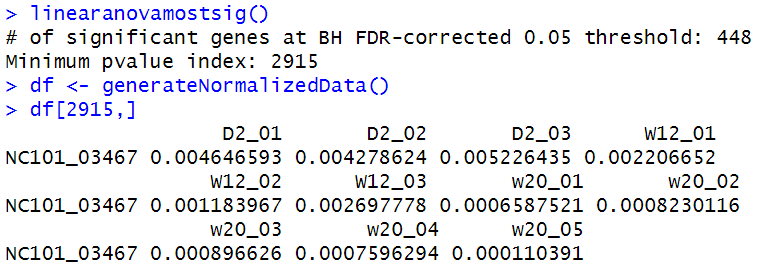
Boxplots for (A), (B) and (C) are at figures 4, 5 & 6, respectively.

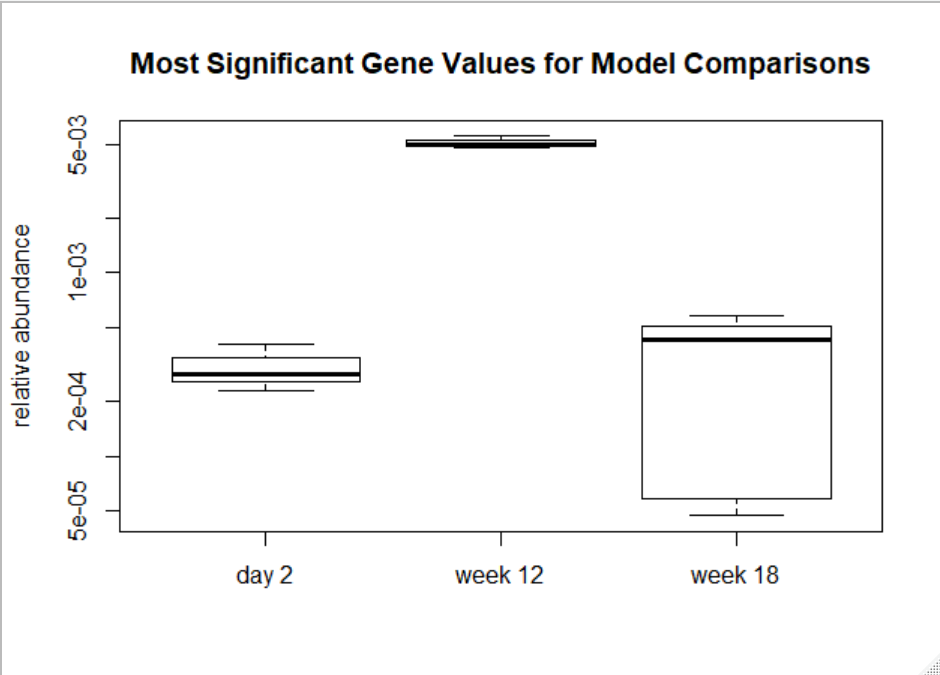


**Figure 4: Most significant gene relative abundance values for one-way ANOVA (full model)**



**Figure 5: Most significant gene relative abundance values for one-way ANOVA as a linear regression (reduced model)**

****



**Figure 6: Most significant gene relative abundance values for ANOVA full versus reduced model comparisons**

1. Overall, do you think the three parameter model in (A) or the two-parameter model in (B) is more appropriate for these data? Justify your answer.