Response Variable Data Processing Protocol

Written by Harmony Martell December 15 2020

These instructions are to process the response variable data for each observation (row, treatment x time) in the hormesis meta-analysis database.

## STEP ONE: **Raw data → Formatted Response Variable Datafiles**

1. Before doing anything, create a duplicate of all the raw response variable datafiles we have received and place them in the folder called /RawRespVarFiles on the Dropbox.
2. Open the raw datafiles that have come from a study and examine them. For example, Study0066 Bellworthy and Fine, 2017 provided two response variable datafiles: [Bellworthy\_2017\_Symb\_Chl\_FvFm.xlsx](https://www.dropbox.com/scl/fi/6mbgeml3dyuwa8klg3rsy/S0066_Bellworthy_2017_Symb_Chl_FvFm.xlsx?dl=0&rlkey=ysosd36jg563e5f6cda002ste) and [Bellworthy\_2017\_Symb\_Chl\_FvFm\_details.xlsx](https://www.dropbox.com/scl/fi/e69p1pfmdd93kq81lxykv/S0066_Bellworthy_2017_symb_chl_fvfm_details.xlsx?dl=0&rlkey=j7pb710rbyr90tm7k39tmb7uq). In each, they have provided details about the total number of treatments in the study (5) and the total number of time points (5). Therefore, there will be 25 observations for this study. This should correspond to the number of rows (i.e., observations) for the study in the database. Check that the number of observations for each species is equal to the trt\_no in **column Z**. If the numbers agree,put an **X** in **column DA** and move on to 3 below. If they don’t, there is a problem in the database and the study must be redone before any additional work can be performed with these data.
3. Open the file called **formattedRespVarDatafileTemplate.xlsx**. For each raw response variable value we receive, we must correctly assign it to the observation number in the database, which represents a single observation from a treatment at a specific point in time (i.e., treatment x time point). Thus, the first task is to organize all response variable values from each study by observation using the **formattedRespVarDatafileTemplate.xlsx**. Please see the steps below to format the data correctly:
   1. Find the observation number assigned to the treatment at the given time point from the study in the database. This can be found in **column D** of the database. Every study will have at least one, but likely many more observations, with unique numbers. ***An observation is defined as all the values from a single treatment at a single point in time***. Treatments that have been sampled multiple times will consist of different observations. For example, in S0066, Bellworthy and Fine, 2017, the authors measured Fv/Fm at five times in all treatments, and there were five treatments, so there are 25 (5 x 5) observations for Fv/Fm. This does not mean there are only 20 measurements. On the contrary, there are multiple organisms in each treatment, and a measurement may be made several times on the same organism. This is considered the technical replicate, and these values will be averaged together using the script.

## STEP TWO: **Formatted Response Variable Datafiles → Database Calculations & Stats**

1. Move all files and the respvarProcess.R script to a single directory on your local machine.
2. Follow each of the steps below to execute the script. The script returns two files, a file called S####\_respvar\_stats.csv that includes summary statistics of each observation and a file called S####\_respvar\_calcs.csv, that gives only the mean observation values needed to populate the database.
3. Open the script respvarProcess.R in R or R studio. Load in the necessary libraries by running Lines 10-12. Set your working directory on Line 15, by specifying the path to the directory the script and datafiles are stored in.

#### A. LOAD THE DATA

Specify the formatted filename for the study you are working on. It should begin with the study number (S####) and followed by the first author name, year and respvar.csv. See the example below:

S0066\_Bellworthy\_2017\_respvar.csv

Examine that the data loaded correctly, but using the head() str() and names() functions (Lines 20-22). Make sure the data have imported with all the correct columns and rows but executing Line 24 (you can open the .csv file in Excel to compare what you see in R to that). Line 25 summary will tell you if you have any missing values. If you should have data in a column and there are NAs, the import has failed. Revisit the .csv file to make sure there are not any errors, and reload the data again.

Lines 26 and 27 create additional response variable columns based on values that have been imported. Head the data again (Line 28) to see that you have all the columns you should.

#### B. PLOT THE DATA BY OBSERVATION

Lines 31 - 38 will plot any of the response variables by observation if they are present. No plot will be generated if the response variable was not provided.

#### C. CALCULATE SUMMARY DATA STATS FOR EACH RESPONSE VARIABLE

1. Symbiont Density (corresponds to Column BZ)
2. Total Chlorophyll (corresponds to Column CD)
3. Chlorophyll a (corresponds to Column CE)
4. Chlorophyll c2 (corresponds to Column CF)
5. Total Chlorophyll per Cell (corresponds to Column CI)
6. Chlorophyll a per Cell (corresponds to Column CJ)
7. Fv/Fm (corresponds to Column CN)
8. Reactive Oxygen Species (corresponds to Column CS)

#### D. COMBINE ALL STATS FOR SAVING

Line 147 combines all the summary statistics for all response variables into a single dataframe.

#### E. SPECIFY A FILENAME AND WRITE THE STATS TO A FILE

***Change the filename*** to reflect the study number (S####) you are processing on Line 153. Failure to do this will overwrite the previous file you were working on! Run Line 153 after you’ve changed the filename and check that it has saved in your working directory.

#### F. COMBINE ALL OBSERVATION MEANS FOR DATABASE

Run LInes 156-166 to create a single dataframe that contains only the mean values of all provided response variables for each observation (i.e., row of the database). You may see only a portion of the dataframe printed to the screen. You can check the variable database\_values using the viewer, or click on it in the Global Environment (R Studio).

#### G. SPECIFY A FILENAME AND WRITE THE STATS TO A FILE

***Change the filename*** to reflect the study number (S####) you are processing on Line 171. Failure to do this will overwrite the previous file you were working on! Run Line 171 after you’ve changed the filename and check that it has saved in your working directory. Open the .csv and add the values of each column to the corresponding column in the database. You can see them below as a reminder:

symbiont density = column BZ

total chlorophyll = column CD

chlorophyll a = column CE

chlorophyll c2 = column CF

total chl per cell = column CI

chl a per cell = column CJ

fv/fm = column CN

ros = column CS