**EXCEL MACROS: DATA PREPARATION & ANALYSIS**

Adolescent Dynamics Lab – Macro Descriptions

August 2015

**Table of Contents**

[1. Altering File Name or File Type 4](#_Toc428449908)

[Adding Filename Ending 4](#_Toc428449909)

[Removing Filename Beginning or Ending 4](#_Toc428449910)

[Resave Excel As New Format 4](#_Toc428449911)

[2. Time Series Preparation 4](#_Toc428449912)

[Adding Task Variables 4](#_Toc428449913)

[Customized Merging 4](#_Toc428449914)

[Logged Variable Creation 4](#_Toc428449915)

[Missing Value Insertion, Removal, Replacement 5](#_Toc428449916)

[Onset Correction 5](#_Toc428449917)

[Removing Excess Time Series Data 5](#_Toc428449918)

[Renaming Columns 5](#_Toc428449919)

[Restricting Range of Time Series Data 5](#_Toc428449920)

[Rounding Data 5](#_Toc428449921)

[Creating Binned Data 6](#_Toc428449922)

[Categorization 6](#_Toc428449923)

[3. Appending 7](#_Toc428449924)

[Adding Columns of Variables with Constant Values 7](#_Toc428449925)

[Adding Participant ID Column 7](#_Toc428449926)

[Append Full Time Series 7](#_Toc428449927)

[Remove Empty Row 7](#_Toc428449928)

[Stacking Variables and Dummy Coding 7](#_Toc428449929)

[4. Creating New Variables 7](#_Toc428449930)

[Baseline Deviations 7](#_Toc428449931)

[Difference Scores 7](#_Toc428449932)

[Lagged Variable Creation 7](#_Toc428449933)

[Logged Variable Creation 7](#_Toc428449934)

[5. Graphing 8](#_Toc428449935)

[Single Channel with Specified Axes Values 8](#_Toc428449936)

[Single Channel with Dynamic Axes and Task Coding 8](#_Toc428449937)

[Two Channels with Specified Axes Values 8](#_Toc428449938)

[6. Correlations 9](#_Toc428449939)

[Within-Subject Zero-Order Correlations 9](#_Toc428449940)

[Customizable Lagged Correlations 9](#_Toc428449941)

[Windowed Correlation Time Series 9](#_Toc428449942)

[7. Noldus Observer Event Logs to TRJ Conversion 9](#_Toc428449943)

|  |  |
| --- | --- |
| 1. Altering File Name or File Type | |
| *These macros do not alter any data within each file, only the name or format.* | |
| Adding Filename Ending | Adds text to the end of all filenames in a folder based on user input. For example, adds “\_HR” to all heart rate time series files. This macro was designed to solve the problem encountered when trying to merge two files of the same name (i.e. if both the HR and GSR files are called “AB123”). |
| Removing Filename Beginning or Ending | Extracts the participant ID (or any desired set of characters) from within a filename. User can remove characters from a filename beginning, ending, or both simultaneously. This macro was designed to solve the problem encountered when trying to merge two files that do not both have the participant ID at the beginning of the filename (i.e. trying to merge AB123\_HR and ResampledAB123\_RSA). |
| Resave Excel As New Format | Opens each excel file and either resaves as a text file in a new folder *or* exports it as a PDF and saves in a new folder. |

|  |  |
| --- | --- |
| 2. Time Series Preparation | |
| *These macros are used for editing and combining the time series files for any variable(s).* | |
| Adding Task Variables | Adds a column containing numeric values that specify which part of an experiment is being completed at each second of a time series. For example, if a participant is listening to instructions for the first 15 seconds of a time series, this macro will create a new “Task” column and assign 0’s to the first 15 rows. The values will then switch to 1’s, 2’s, etc. as the task changes. These values are calculated using time values input by the user in Sheet 2.  **Note:** Do NOT rename the macro file or the names of Sheet1 and Sheet2 as the code itself would also have to be altered to recognize the new names. |
| Customized Merging | Combines the data from multiple different time series by copying from one file and pasting into another. For example, copies the GSR data from one participant and pastes it into a new column within that same participants HR time series. This macro ensures the participant ID is the same for the two files being merged, and adds missing values if a participant is missing one of the two files (i.e. has GSR time series but no HR).  **Note:** Prior to merging, all time series should have their onset corrected. |
| Logged Variable Creation | Takes the log of all values in a specified column and creates a new column containing the new logged values. This macro was designed to get more useful information from the RSA time series. |
|  |  |
|  |  |

|  |  |
| --- | --- |
| **2. Time Series Preparation (Continued)** | |
| Missing Value Insertion, Removal, Replacement | This file contains four slightly different macros that work with missing values.  **Insert Missing:** Replaces all blank cells with missing values. **Remove Missing:** Replaces all missing values with blank cells. **Replace Missing:** Changes all missing values to a user-specified value. **Replace with Missing:** Replaces all occurrences of a user-specified value (within a specified range) with missing values. This macro will only search the specified columns (i.e. if you want to replace all 0’s with missing values, can ensure the 0 in the onset column is excluded). |
| Onset Correction | Inserts a time value of 0 (representing the *onset* of the physiological response) and shifts all physiological data up by one cell. The final row in the time series will therefore be blank. This macro is generally the first step when editing time series files. |
| Removing Excess Time Series Data | Shortens time series files of varying length to match the desired lengths specified by the user. This macro is used when the total length of the time series file is different for each participant (for example, when the instructions before a task are included in the time series). The Acknowledge scripts used to obtain the time series files are run for all participants using the longest total time, therefore the excess data must be deleted afterwards. The user must input the participant ID and total time values for each participant in Sheet2.  **Important:** Time series must already have corrected onset!  **Note:** Do NOT rename the macro file or the names of Sheet1 and Sheet2 as the code itself would also have to be altered to recognize the new names. |
| Renaming Columns | Changes the names of all columns within a time series. Files in the input folder may contain up to 15 columns in total.  **Note:** This macro saves over the original file. |
| Restricting Range of Time Series Data | Restricts the range of values in a time series by altering any numbers below a specified minimum and above a specified maximum. All values below the specified minimum will be replaced with that desired minimum value. All values above the specified maximum will be replaced with that desired maximum value. All other values will be unaffected. |
| Rounding Data | Rounds all values within a specified column (or range of columns) to a specified number of decimal places. |

|  |  |
| --- | --- |
| **2. Time Series Preparation (Continued)** | |
| Creating Binned Data | Creates a time series of binned values by taking the weighted average across each “bin” of a user-specified size. For example, if user enters a bin size of 30 seconds, the original time series will be divided into 30-second segments and a weighted average will be calculated for each segment. The resulting file will still contain an onset column and one or more data columns; however, the onset column will now contain only multiples of 30 and the values in each data column will correspond to those 30-second weighted averages. The resulting files are saved within a new folder, created as a subfolder of the input folder.  If the file length itself is not an exact multiple of the bin size you select, the last bin in the data set will be incomplete. The user has the option of including or excluding this final partial bin. |
| Categorization | Categorizes continuous time series data into intervals of equal width to allow for later categorical analysis (generating State Space Grids, for example). The user specifies the desired number of categories they wish to create for each variable. In order to create these categories, the macro must first calculate the maximum and minimum value of each variable, then determine the difference and divide by the desired number of categories. Each data point within the user-specified column(s) is then assigned to a category, and this category value is pasted as a range in the first available column.  The user has the option of using the same max, min and categories for **all** participants (i.e., macro finds the *overall* max and min and uses these for creating categories) or using a unique max, min and set of categories for each participant (i.e., macro finds the specific max and min for each file and uses these to create categories).  The user also has the option of including a column of category labels, where the column adjacent to the newly created categorized column is filled with integers representing where each category falls with respect to the other categories. For example, if “-10 to -2” is the first or lowest category, the column next to it would receive a “1”.  **Note:** Original files must be in percent change format, where values are either decimals or percentages representing the deviation of each value from a baseline value. In order to get data into this format, run the [“Baseline Deviations”](#_Baseline_Deviations) macro. |

|  |  |
| --- | --- |
| 3. Appending *These macros are used to create appended files for multilevel modeling analyses. The resulting appended files include all participants stacked within a single Excel spreadsheet.* | |
| Adding Columns of Variables with Constant Values | This macro adds a time-invariant variable (i.e. a variables whose value does not change throughout the time series [e.g. age, sex, etc.]). Data for all variables of interest must be entered in Sheet2 of the worksheet. The macro will match the Participant ID in the filename to the Participant ID in column A of Sheet2 and input the corresponding variable value. |
| Adding Participant ID Column | This macro isolates the Participant ID for each file and creates a PartID column. Onset values remain in Column 1, PartID is inserted in Column 2, and all other columns are shifted one column to the right.  Each filename within the input folder **must begin with** the participant ID. If this is not the case, run the [“Removing Filename Beginning”](#_Removing_Filename_Beginning) macro. |
| Append Full Time Series | Creates an appended file for data containing time series variables (i.e. heart rate, GSR) as well as invariant variables (i.e. age, sex, BMI). If the time series is 180s long, the participant ID and invariant variable data is repeated for all 180 rows. The onset and time series variables will be different for each row within the series. |
| Remove Empty Row | This macro will clear any data in and beyond a user specified row (i.e. the desired last row of data) |
| Stacking Variables and Dummy Coding | This macro will append each variable in a data file to the bottom of the previous variable. The macro will then dummy code to indicate which variable is shown. It assumes that the first column in each file contains onset values and that each file has a column containing participant ID. |

|  |  |
| --- | --- |
| 4. Creating New Variables | |
| *These macros create files with* new *variables that can then be used for specific types of analyses.* | |
| Baseline Deviations | Compares the entire time series from a task to a single average baseline value. For example, compares each second-by-second HR value during a speech task to the average HR during a baseline task. The resulting file contains a “time series” of deviation scores. |
| Difference Scores | Creates a series of within-subject difference scores by subtracting the lag variable at each second from the 0-onset variable.  **Note:** Lagged variable must be created first and added to a merged file containing the 0-onset variable (this can be done using the [“Customized Merging”](#_Customized_Merging) macro) |
| Lagged Variable Creation | Creates lagged variables by offsetting the data in a variable column by a specified number of seconds. |
| Logged Variable Creation | Takes the log of all values in a specified column and creates a new column containing the new logged values. This macro was designed to get more useful information from the RSA time series. |

|  |  |
| --- | --- |
| 5. Graphing | |
| *These macros use time series data to create graphical representations.* | |
| Single Channel with Specified Axes Values | Creates a smooth scatterplot using onset along the x-axis and a user-specified variable along the y-axis. User can customize the axes by inputting their desired minimum and maximum values. The resulting graphs are copied and pasted into a Word document, and will have the same user-specified x- and y-axis scales for all participants. |
| Single Channel with Dynamic Axes and Task Coding | Creates a smooth scatterplot using one user-specified variable along the x-axis and another user-specified variable along the y-axis. User does **not** indicate a desired scale for the axes. Instead, axes will be formatted to fit the minimum and maximum values of each participant. The resulting graphs are copied and pasted into a Word document and will have **different** x- and y-axis scales for each participant.  This macro also gives users the option to colour-code the graph using the previously mentioned task variables. If this option is selected, the resulting graph will included segments of different colours, corresponding to the different parts of the task. For example, the points corresponding to “Instructions” will be one colour, and the points corresponding to “Speech” will be a different colour. This macro is able to code up to 5 different colours in a single graph.  **Note:** In order to use the colour-coding feature, the [Adding Task Variables](#_Adding_Task_Variables) macro must be run first. |
| Two Channels with Specified Axes Values | This macro graphs two time series variables (e.g. RSA and GSR) along the y-axis against “onset” along the x-axis. The range of both axes is specified by the user and will therefore be the same for each graph. The legend titles on each graph are taken from the column headings of each corresponding variable. The graphs are pasted into a single word document. The last row of data can be excluded if it contains missing values. |

|  |  |
| --- | --- |
| 6. Correlations | |
| *These macros calculate various correlations between two physiological variables.*  ***Note:*** *Variables to be correlated must be merged into a single file first.* | |
| Within-Subject Zero-Order Correlations | Creates straight-forward correlations between two user-specified columns within a merged file. Pastes the correlation R2 value and the Concordance Correlation Coefficient (CCC) in the output sheet. |
| Customizable Lagged Correlations | User specifies a maximum lag. The macro will then calculate the appropriate number of correlations in both the positive and negative direction, as well as a correlation with a lag of 0.  For example, if the user specifies a lag of 5 seconds, the macro generates a total of 11 correlation coefficients for each participant: no lag (zero seconds), positive lag (1 to 5 seconds), and negative lag (-1 to -5 seconds). The macro then determines which of the 11 values is the largest and pastes it into the output sheet with its corresponding lag time. For example, if the largest correlation was 0.5678 and was calculated at a lag of +3 seconds, then “0.5678” is pasted to the output sheet and “3” is pasted in the adjacent cell.  The chosen correlation value is then used to calculate an R2 value, which is then pasted in the next available column. |
| Windowed Correlation Time Series | Calculates a series of correlation coefficients within windows of a user-specified length. For example, a 20-second window begins by calculating the correlation for 0-20 seconds, then shifts to 1-21 seconds, then 2-22 seconds, etc. |

|  |
| --- |
| 7. Noldus Observer Event Logs to TRJ Conversion |
| Cycles through the event logs of all participants and creates a text file to indicate the respective behaviours of a mother daughter dyad, for example, at a given time point. Resulting file contains 3 new columns: onset, mother behaviour, and daughter behaviour.  **Note:** The time column contains only the times at which behaviours were observed, not a complete time series. |