READ ME: Drug Poisoning Indicators WG TREND ANALYSIS SAS Program

Last Updated 10/12/18 By Hannah Yang (Hannah.yang@mt.gov)

The Trend Analysis SAS code has 6 general steps. MACROs are used to repeat each step for both ED Encounter (ED) and Hospital discharge (HOSP) datasets:

1. Restrict dataset to specified analysis date range, apply inclusion/exclusion criteria according to the CDC Opioid Overdose Indicator Support Toolkit v3.0, convert ICD codes to proper format
2. Identify drug poisoning cases according to Toolkit v3.0 definitions (any mention approach) & restrict datasets to cases only
3. Export monthly counts for H1, H2, H3 by age group
4. Decompose the original time series into trend-cycle component (TCC), seasonal component, and irregular component. Add variables for pre-intervention, intervention, post-intervention, & time
5. Using only pre-transition TCC data points, forecast the post-transition TCC values of H1, H2, H3
6. Fit Interrupted Time Series model to estimate level and slope change in H1, H2, H3

INTRODUCTORY

The program includes a few lines of introductory code, explained below:

```
OPTIONS SYMBOLGEN MPRINT MLOGIC;
```

- These are MACRO debugging options. Each adds different information to the SAS log file, which is useful if you are curious about how the MACRO variables resolve. If you do not wish to see such information in your log file, change to NOSYMBOLGEN, NOMPRINT, NOMLOGIC

```
%LET TYPE1=ED;
%LET TYPE2=HOSP;
```

- These MACRO variables allow for the use of iterative MACRO statements (`%DO Q=1 %TO 2;`) and &&var&i variables (double ampersand MACRO variables) to loop through ED and HOSP datasets

Example code from the Validation Dataset has been included throughout the program. Where this is denoted, users should replace it with their own code.

CUSTOMIZATION

The first block of code contains a numbered list of MACRO variables defined with %LET statements. Users should assign custom values to each of the MACRO variables reflecting the desired analysis dates and the specific characteristics of their jurisdiction's dataset. Example values from the Validation Dataset are included initially.

The subsequent code references the MACRO variables rather than jurisdiction-specific variable names, which improves generalizability of the SAS code while greatly reducing the effort required to read & run it. If the value of a MACRO variable is "none" or "not applicable" for your jurisdiction, assign it a blank value by submitting %STR();

Note that is no separate program for administrative datasets with or without dedicated field(s) for external cause codes. If your jurisdiction does NOT have dedicated external cause code field(s), simply assign a blank value to the MACRO variable containing the names of variables for dedicated external cause code fields:

```
%LET EC_VARS=%STR();
```

Step 0: Bring your datasets into SAS work library

- Define the location on your hard drive of your original ED and HOSP datasets using LIBNAME statement
- Modify the SET statements according to your jurisdiction to bring your ED and HOSP datasets into SAS work library. Examples from the Validation Dataset are included initially.

Step 1: Clean and format the datasets

- Restrict to desired analysis dates
- Apply inclusion/exclusion criteria:
    - Include only in-state residents
    - Include only non-federal, acute care, or inpatient facilities (exclude VA and other federal hospitals, rehabilitation centers, psychiatric hospitals)
    - Exclude in-hospital deaths and deaths in ED
    - If necessary, exclude admitted patients from ED Dataset
- Convert ICD10CM codes in the diagnosis and external cause fields into the proper format
- Create the following variables:

Variable Name	**Description**
ICD_ERA	ICD9CM or 10CM Era (”9” or “10”)
DISYEAR	Discharge year (Calendar Year)
DISMO_BEGINDATE	Discharge year and month (YYYY-MM) (Calendar Year)
DISQTR_BEGINDATE	Discharge year and quarter (YYYY-Q) (Calendar Year)
FFY_DISYEAR	Discharge year (Federal Fiscal Year)
AGEG	Age group
STATE	Name of state or jurisdiction
PTTYPE	=“ED” for ED dataset or “HOSP” for HDD dataset

Step 2: Identify Drug Poisoning Cases

- The indicator definitions are based on the CDC’s Opioid Overdose Indicator Support Toolkit v3.0. Refer to the toolkit for ED Encounter and Hospitalization specific definitions of the three indicators for both ICD9CM and ICD10CM:
    - H1: All drug overdose
    - H2: Opioid overdose excluding heroin
    - H3: Heroin overdose
- Due to the nature of the ICD9CM and ICD10CM definitions for the 3 drug overdose indicators, different approaches are employed to identify drug poisoning cases:
    - For ICD9CM records, See MACRO %ICD9_TAG_CDCv3
        - PRXMATCH() function is used to search for records with a Principal/First Listed Diagnosis of H1, H2, or H3
        - An Array is used to search the external cause fields (if applicable) and diagnosis fields for any mention of an H1, H2, or H3 E-code

```
ARRAY ICD9_ECODE_ARRAY{*} $ &EC_VARS &PRDX &SUBDX_VARS;
```

- For ICD10CM records, See MACRO %ICD10_TAG_CDCv3
  - An array is used to search the diagnosis fields and external cause fields (if applicable) for any mention of an H1, H2, or H3 Diagnosis code

```
ARRAY ICD10_DXCODE_ARRAY{*} $ &PRDX &SUBDX_VARS &EC_VARS;
```

- Note that records containing a Heroin poisoning diagnosis code or E-code (H3) are excluded from the H2 indicator to avoid double counting of records with multiple types of opioids. See below for examples of when a single record indicates presence of both heroin and non-heroin opioids:
  - For ICD9CM: H1=1, **H2=0**, H3=1, ICD9_H1EC=1, ICD9_H2EC=1, ICD9_H3EC=1, ICD9_H1_ECODE= "E8500", ICD9_H2_ECODE= "E8502", ICD9_H3_ECODE= "E8500"
  - For ICD10CM: H1=1, **H2=0**, H3=1, ICD10_H1_DXCODE= "T400X1A", ICD10_H2_DXCODE= "T400X1A", ICD10_H3_DXCODE= "T401X1A", ICD10_H1_DEPTH=1 ICD10_H2_DEPTH=1 and ICD10_H3_DEPTH=3

- Create the following variables:

Variable Name	Description
H1	=1 IF ALL DRUG OVERDOSE CASE
H2	=1 IF OPIOID OVERDOSE EXCLUDING HEROIN CASE
H3	=1 IF HEROIN OVERDOSE CASE
ICD9_H1DX	=1 IF PRINCIPAL/1ST LISTED DX IS H1 DIAGNOSIS CODE (ICD9CM)
ICD9_H2DX	=1 IF PRINCIPAL/1ST LISTED DX IS H2 DIAGNOSIS CODE (ICD9CM)
ICD9_H3DX	=1 IF PRINCIPAL/1ST LISTED DX IS H3 DIAGNOSIS CODE (ICD9CM)
ICD9_H1EC	=1 IF ANY MENTION OF H1 E-CODE IS FOUND (ICD9CM)
ICD9_H2EC	=1 IF ANY MENTION OF H2 E-CODE IS FOUND (ICD9CM)
ICD9_H3EC	=1 IF ANY MENTION OF H3 E-CODE IS FOUND (ICD9CM)
ICD9_H1_ECODE	H1 E-CODE (ICD9CM)
ICD9_H2_ECODE	H2 E-CODE (ICD9CM)
ICD9_H3_ECODE	H3 E-CODE (ICD9CM)
ICD10_H1_DXCODE	H1 DIAGNOSIS CODE (ICD10CM)
ICD10_H2_DXCODE	H2 DIAGNOSIS CODE (ICD10CM)
ICD10_H3_DXCODE	H3 DIAGNOSIS CODE (ICD10CM)
ICD10_H1_DEPTH	POSITION OF H1 DX CODE WITHIN ARRAY (ICD10CM)
ICD10_H2_DEPTH	POSITION OF H2 DX CODE WITHIN ARRAY (ICD10CM)
ICD10_H3_DEPTH	POSITION OF H3 DX CODE WITHIN ARRAY (ICD10CM)

Step 3: Export monthly counts for H1, H2, H3

- Define the location on your hard drive of where you will save all your exported results using the %LET statement
- Create CLASS dataset to ensure the excel sheets are standardized between all jurisdictions (ensures that all categories are included even if there are no observations)
- Export Monthly counts for H1, H2, H3 by Age Group
- Creates 1 EXCEL Workbook: *YOURSTATE_Monthly_DrugCounts_2010-2016.xlsx*

Step 4: Decomposition

- Classical Decomposition is used to separate the original time series data into trend-cycle component (TCC), seasonal component, and irregular component.
- The objective is to remove the effects of seasonality and look directly at the TCC to evaluate whether there are any discontinuities in the temporal pattern. In other words, did the ICD10CM transition affect drug poisoning trends in the absence of seasonal effects?
- PROC TIMESERIES creates 2 datasets each for ED and HOSP data
    - **ED_TS**, **HOSP_TS**: Time series datasets with 1 row per time period/bin (YYYY-MM) and columns containing monthly counts for H1, H2, H3
    - **ED_DECOMP, HOSP_DECOMP**: Decomposed time series datasets with 1 row per indicator per time period (H1 YYYY-MM). Columns include the original series value (monthly count), seasonal index (1-12), TCC (2x12 moving average), seasonal component, & irregular component. Dummy variables are added:
        - PRESLOPE: Coded sequentially starting from 1 for the pre-transition time points (ICD9CM Era), staying constant after the transition (ICD10CM Era). Models the structural trend (slope) before the transition.
        - INTERVENTION: Coded 0 for pre-transition (ICD9CM Era) time points, and 1 for post-transition (ICD10CM Era) time points. Models the level change between time points immediately before and after the transition.
        - POSTSLOPE: Coded 0 for pre-transition time points (ICD9CM Era), and sequentially from 1 after the transition (ICD10CM Era). Models the structural trend (slope) after the transition
        - TIME: Coded sequentially starting from 1 for the entire study period
- Creates 2 HTML Outputs: *YOURSTATE_ED_DECOMPOSITION.html, YOURSTATE_HOSP_DECOMPOSITION.html*
- These will include graphs of the original series and decomposition components for each indicator H1, H2, H3
- For more information on decomposition, see the resources below:
    - https://anomaly.io/seasonal-trend-decomposition-in-r/
    - https://www.otexts.org/fpp/6/3

Step 5: Forecasting the counterfactual scenario using Trend-Cycle-Component (TCC)

- Create forecast dataset with empty data points for post-transition TCC values: **ED_FORECAST, HOSP_FORECAST, ED_FORECAST_GRAPH, HOSP_FORECAST_GRAPH**
- Test pre-transition TCC Data for Autocorrelation (see below)
- Using only the pre-transition TCC data points, predict post-transition TCC values for H1, H2, H3 using PROC AUTOREG
    - Creates 2 HTML outputs: *YOURSTATE_ED_forecastMODEL.html, YOURSTATE_HOSP_forecastMODEL.html*
- Graph Forecasted vs. Observed trends for H1, H2, H3

STEP 5.1- Test pre-transition TCC Data for autocorrelation

- One of the assumptions when building a linear regression model is that the errors are independent. In time series data, error terms may be correlated over time- autocorrelation
- To test for autocorrelation, use the Durbin-Watson test. This is done by running PROC AUTOREG with pre-transition TCC data (TCC_PRE) as the response variable and time (TIME) as the predictor, with DW=12 for monthly data and DWPROB option to display all the Durbin-Watson statistic p-values. If any of the 12 p-values are <0.05 it indicates presence of autocorrelation
    - Creates 2 HTML outputs: *YOURSTATE_ED_forecastAUTOCORRTEST.html, YOURSTATE_HOSP_forecastAUTOCORRTEST.html*
    - These will contain Durbin-Watson statistics and p-values, Plot of residuals over time (RES), Predicted and actual values over time (FITPLOT), ACF versus Lag (ACF), PACF versus lag (PACF)

STEP 5.2- Backwards Stepwise Regression (If autocorrelation is detected)

- Durbin-Watson test indicates the presence of autocorrelation, but does not provide information about the order of autocorrelation. 'Order' is the number of immediately preceding values in the series that are used to predict the value at the present time
- Use backwards stepwise regression to look at 'order'. This is done by running PROC AUTOREG with pre-transition TCC data (TCC_PRE) as the response variable and time (TIME) as the predictor, with the BACKSTEP option and NLAG=13 for monthly data
    - Creates 2 HTML outputs: *YOURSTATE_ED_forecastSTEPWISE.html, YOURSTATE_HOSP_forecastSTEPWISE.html*
    - These will show which autoregressive parameters were insignificant and eliminated, and which were retained as significant (use SLSTAY= option to control significance level).

STEP 5.3- Determine order of autocorrelation

- Considering both the stepwise regression results and PACF plot, determine the order of autocorrelation for each indicator for ED and HOSP data.
    - On the backwards stepwise regression results, look at the "Estimates of Autoregressive Parameters" report to see which autoregressive parameters were retained. These are possible choices for the order of an autoregressive model
    - On the PACF plot produced from the Durbin-Watson test, look for large PACF values (outside of the two standard error bands) at a given non-zero Lag to indicate possible choices for the order of an autoregressive model
- Enter the 'order' of your models in the MACRO variables using %LET statements. If no autocorrelation is detected, simply assign a blank value to the MACRO variable, for example:

```
%let ED_forecastORDER_H1=();
```

- There will be 6 models in total, and each may have a different order of autocorrelation (or none). These parameters will be included in the Forecasting Model, for example:

```
%let ED_forecastORDER_H1=(1);
%let ED_forecastORDER_H2=(1);
%let ED_forecastORDER_H3=(1 2);
%let HOSP_forecastORDER_H1=(1 2);
%let HOSP_forecastORDER_H2=(1);
%let HOSP_forecastORDER_H3=(1 2);
```

- For more information on autocorrelation, see:
    - http://support.sas.com/documentation/cdl/en/etsug/63939/HTML/default/viewer.htm#etsug_autoreg_sect002.htm

Step 6: Fit Interrupted Time Series Model using Trend-Cycle-Component (TCC)

- Test TCC Data for Autocorrelation (see below)
- Fit ITS model for H1, H2, H3 using PROC AUTOREG
    - Creates datasets **ED_ITS, HOSP_ITS, ED_ITS_GRAPH, HOSP_ITS_GRAPH**
    - Creates 2 HTML outputs: *YOURSTATE_ED_itsMODEL.html, YOURSTATE_HOSP_itsMODEL.html*
    - These will contain the "Autoregressive Parameters Assumed Given" Report with parameter estimates of the coefficients for predictor variables:
        - Coefficient for PRESLOPE represents the average monthly change in H1, H2, H3 before the transition to ICD10CM
        - Coefficient for INTERVENTION represents the immediate effect of the transition, or the change in level (drop or jump) immediately following the transition
        - Coefficient for POSTSLOPE represents the average monthly change in H1, H2, H3 after the transition to ICD10CM
- Graph Interrupted Time Series model vs observed for H1, H2, H3

STEP 6.1- Test pre-transition TCC Data for autocorrelation

- Use the Durbin-Watson test by running PROC AUTOREG with TCC as the response variable and PRESLOPE, INTERVENTION, and POSTSLOPE as predictor variables, with DW=12 for monthly data and DWPROB option
    - Creates 2 HTML outputs: *YOURSTATE_ED_itsAUTOCORRTEST.html, YOURSTATE_HOSP_itsAUTOCORRTEST.html*

STEP 6.2- Backwards Stepwise Regression (If autocorrelation is detected)

- Use backwards stepwise regression to look at 'order' by running PROC AUTOREG with TCC as the response variable and PRESLOPE, INTERVENTION, and POSTSLOPE as predictor variables, with the BACKSTEP option and NLAG=13 for monthly data
    - Creates 2 HTML outputs: *YOURSTATE_ED_itsSTEPWISE.html, YOURSTATE_HOSP_itsSTEPWISE.html*

STEP 6.3- Determine order of autocorrelation

- Enter the 'order' of your models in the MACRO variables using %LET statements. There will be 6 models in total, and each may have a different order of autocorrelation (or none). These parameters will be included in the ITS Model, for example"

```
%let ED_itsORDER_H1=(1);
%let ED_itsORDER_H2=(1);
%let ED_itsORDER_H3=(1 2);
%let HOSP_itsORDER_H1=(1 2);
%let HOSP_itsORDER_H2=(1);
%let HOSP_itsORDER_H3=(1 2);
```