# Central Tendency White Paper Requirements Specification

## Scripts for Figure 7.2 Box Plot – Change in xxx Over Time

### Script Specification

### Assumptions:

* Input dataset assumption – Endpoint timepoint is already in dataset.

#### Specific Output Requirements

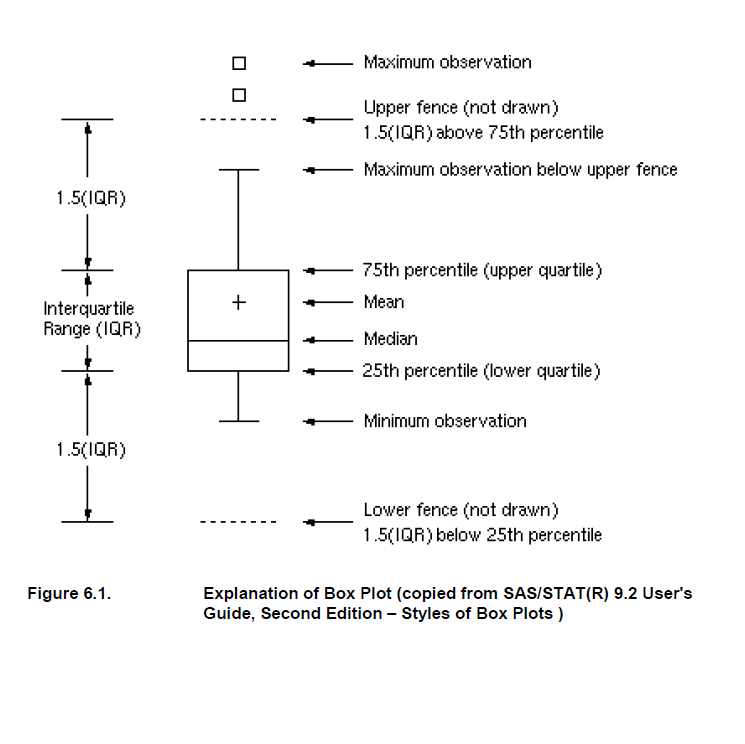
* Generates a box plot of CHG by AVISIT (optionally ATPT as well) and TRTP/TRTA. See domain variations in Usage Requirements, below
* Basic script functionality (user options)
* ***data display features***
* Schematic or Tukey notched box plots, as described in [SAS/STAT 9.4 Graph Template Language: Reference (Boxplot Statement)](http://support.sas.com/documentation/cdl/en/grstatgraph/67882/HTML/default/p0vuh82v39fsasn1vqhzmhdl8y16.htm#n0g9wrnpg9zsryn1dk5csdrdfnqy)
* analyze by treatment
* group the boxes by time point, and label each time point clearly
* outliers:
* **IQR outliers** as box outlines, and
* Zero reference line
* ~~Find last populated CHG value for each subject when sort by x-axis variables and treatment. Plot this value in slotted timepoint and as the Endpoint.~~
* ***methodology***
* Change-from-baseline to Endpoint includes all subjects with both a baseline and post-baseline measurement
* P-value comparing Active Treatment with Comparator, using ANCOVA containing terms for treatment and the continuous covariate of baseline measurement
* ***User variability***
* library and name of each input data sets
* output location (user-specified path)
* option to suppress the p-value for Endpoint treatment comparison
* parameterize core variables required for analysis & display:
* "Treatment name" which could be planned or actual
* "Treatment number" used for display order of Treatment Names (TRTAN/TRTPN) are not required. If the numeric is not on the input dataset, sort character treatment variable alphabetically and use that as sort order.
* CHG variable (change-from-baseline outcome)
* "Baseline" visit number for ANOVA comparison & p-value (e.g., an AVISITN value)
* "Endpoint" visit number for ANOVA comparison & p-value (e.g., an AVISITN value)
* SAFFL variable (population flag) – allow for any population flag from ADSL
* ANL01FL variable (analysis flag) – allow for any subsetting (where-clause statement) on input data
* ANCOVA p-value option: Leave these values blank to omit Endpoint ANCOVA p-values
* "Baseline" measurement variable , such as BASE
* "Reference" treatment number, such as a TRTPN value like zero (0) – allow for character value if numeric treatment variable is not on input dataset
* Allow for use of AVISIT or (AVISIT and ATPT as x-axis variables)
* ***basic readability features***
* maximum number of boxes to display on a page (basic paging control)
* Restrictions
* If BASE variable is missing then exclude observation.
* Exclude any unplanned observations (should be accounted for in subsetting where-clause parameter
* Page Break Requirements:  
  Allow parameter for page break – either maximum number of columns on a page or break on a variable
* Footnote:
* Box plot type is schematic: the box shows median and interquartile range (IQR, the box height); the whiskers extend to the minimum and maximum data points within 1.5 IQR of the lower and upper quartiles, respectively. Values outside the whiskers are shown as outliers. Means are marked with a different symbol for each treatment. P-value is for the treatment comparison from ANCOVA model Change = Baseline + Treatment.

#### General Output and Formatting Requirements

See the specification document **CS\_GeneralOutputandFormattingRequirements.docx**.

* Page margins: graphic fits within 9" x 6.5" common area for A4 and Letter paper.
* Titles:
* Centered
* Footnotes:
* Left-justified
* Colophon – Each graphic displays:
* program name,
* output name,
* date generated, and
* location of source data
* P-values:
* For p > 0.1, report the p-value to 2 decimal places
* For 0.1 > p > 0.001, report the p-value to 3 decimal places
* For p < 0.001, report p < 0.001.
* Decimal places for continuous data summaries:
* Mean and estimates of precision (e.g., variance, SD, SE, Confidence Intervals) have 1 more digit than the most-precise collected data.
* Distributional parameters (eg, Q1, Q3) also have 1 more digit.
* Minimum and Maximum have the same precision as the most-precise collected data.
* All diplays summarizing baseline data must include a "total" column for studies with more than 1 arm.

#### Box Plot Elements



### Usage Requirements

#### Required Input

##### BDS Domain

* Dataset: ADxx (ADVS, ADEG, ADLB, ADLBC, ADLBH, ADLBHY, …)
* Variables: STUDYID, USUBJID, SAFFL, TRTP/TRTA, TRTPN/TRTAN (if present), PARAM, PARAMCD, CHG, BASE, AVISIT, AVISITN (if present), ATPT (optional), ATPTN (if present and if ATPT is selected)
* Record selection: Allow for any population subsetting as well as other input dataset subsetting (i.e. SAFFL='Y' and ANL01FL='Y’)

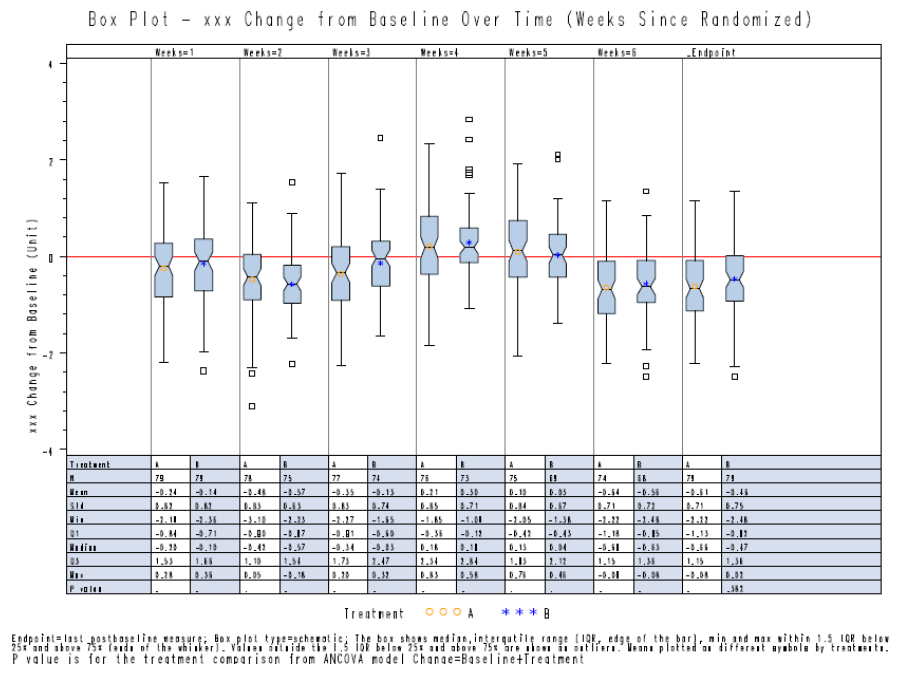
#### Macros

* Requires the PhUSE CS macro library:   
   <https://github.com/phuse-org/phuse-scripts/tree/master/whitepapers/utilities>
* User must ensure that SAS can find PhUSE CS macros in the SASAUTOS path

### Test Data

* Data sets from <https://github.com/phuse-org/phuse-scripts/tree/master/data/adam/cdisc>
* Domain data
  + Vital Signs Domain
  + ADVS – Vital signs
  + ADLBC, ADLBH, ADLBHY – Laboratory measures
  + ??? – ECG measures

### Sample Output



### Reference Documents:

White paper: <http://www.phusewiki.org/wiki/images/4/48/CSS_WhitePaper_CentralTendency_v1.0.pdf>

Programming Guidelines: <http://www.phusewiki.org/wiki/index.php?title=WG5_P02_Programming_Guidelines>