Analysis of Data from DEOCS Survey: Weighting by Decision Trees, Evaluation of Weights, and Non-Response Bias Analysis

DEFENSE EQUAL OPPORTUNITY MANAGEMENT INSTITUTE
DIRECTORATE OF RESEARCH DEVELOPMENT AND STRATEGIC INITIATIVES

Dr. Richard Oliver Hope Human Relations Research Center
Directed by Dr. Daniel P. McDonald, Executive Director
366 Tuskegee Airmen Drive Patrick AFB, FL 32925
321-494-2747

Prepared by
Dr. Wei Wan
Dr. Daniel P. McDonald

DEOMI PRESS

Tech Report No. 20–02

Abstract

The dataset of this study is DEOCS survey for October 2017 to March 2018. Decision Tree method has been adopted to compute weights for respondents. Three types of weights have been computed for SAPR, DSPO, and ODMEO, separately. According to the guideline from OPA for evaluating Non-Response Bias (NRB), we use these three types of weights to conduct NRB analysis by (1) comparing the composition of the sample compared with survey respondents by key demographics; (2) comparing estimates from the NRB follow-up survey. Furthermore, by the technical recommendations from OPA, we evaluate the validity and reliability of the final weights by (1) checking whether all eligible respondents have final weights, and whether all other sample members do not have final weights; (2) checking whether the summation of final weights equals the population totals; (3) checking the trend line of the key survey weighted estimate overall and in key subpopulations to see whether the changes across waves are reasonable; (4) checking the unequal weighting effect (UWE) of the final weights and the design effects of the weighted estimates to see whether there is large weight variability and whether any extremely large weights have an impact on the weighted estimates. Five algorithms have been developed with built-in Python program in SPSS to conduct NRB analysis and evaluation of weights. The major results are: (1) no extreme weights; (2) the NRB ratios for all surveys are minor; (3) the difference of effects between un- and normalized weights is also little.

Table of Contents

List of Tables

List of Figures

1. Introduction

Bias could be introduced by survey nonresponse in the survey estimates. It is a common practice to apply weighting adjustments in surveys to compensate for nonresponse and noncoverage and to make weighted sample estimates conform to population values. Weights cannot correct all types of survey error (questionnaire design, data collection, sampling, nonresponse effect), but we try to make the weighted records represent the population of inference as closely as possible.

In this project we use weights to adjust non-responses in DEOCS survey so the weighted estimates of following survey variables approximate true population values: *Commitment_Fav,SrLeadership_Fav,OrgPerf_Fav,GrpCoh_Fav,Trustlead_Fav,JobSat_Fav,OrgProc_Fav,Eng_Fav, Inclusion_Fav, SexHar_Fav, Prevent_Fav, Response_Fav, SARetaliat_Fav, SHRetaliat_Fav.*

The dataset for this project is: “01. Cumulative with NR weights October 2017 - March 2018_For Dr. Wan.sav”. It contains 612821 cases and 157 variables. It contains four weights:

NonRespWeightCHAID_SAPR

NonRespWeightCHAID_DSPO

NonRespWeightCHAID_ODMEO

NonRespWeightCHAID_TrueIndicator

It is required that when using each of above weights, the data only contains the responders for each report constituency. That is, if using *NonRespWeightCHAID_SAPR*, the dataset only contains the responders with variable *Indicator_SAPR = 1*. Furthermore, we will analyze Nonresponse Bias by studying the effects of above four weights on demographics and above observable characteristics/variables.

2. Type of Missing Data

In any survey research, it is very normal to have missing data. Missing data mainly comes from four sources (Brick & Kalton, 1996):

(1) Total or unit nonresponse. This is the most common source of missing data, which happens when the selected sampled elements do not respond to the survey. In this project, each element refers to a person. The nonresponse results from non-contacts, or simple refusal to participate in the survey. The common practice to compensate for the total nonresponse is by weighting adjustments, where the respondents are assigned weight greater than one in the analysis to represent the non-respondents.

(2) Noncoverage. This source of missing data happens when the survey's sampling frame does not include some types of elements in the population. We are missing those elements because they have no chance of being selected and hence go unrepresented. We can compensate noncoverage by weighting adjustments. However, the weighting reference is different. For total nonresponses, the non-respondents can be identified within the selected sample, so the weighting adjustment can be done by using sample data alone. For noncoverage, we do not have any information about missing elements from the sample, so the weighting reference has to be based on external data sources, such as population data.

(3) Item nonresponse. This source of missing data comes from a sampled element which participates in the survey but fails to provide responses to one or a small number of survey items. This type of nonresponse may come from refusals, not-knowing the answer, inconsistency with answers, or the interviewer failing to ask this question or record the response. The usual method to compensate for item nonresponses is imputation, which assigns a value to the element with missing values.

(4) Partial nonresponse. This source of missing data is similar to total and item non-responses because partial nonresponse means the survey misses a substantial number of item nonresponses. This could happen when a respondent quits the survey in the middle or a respondent in a panel survey fails to participate in one or more of the waves of the panel. We have two methods to compensate for missing data for partial nonresponse. One is weighting. In this approach, we just drop the partial non-respondents from the data file, and apply weighting adjustments used in total nonresponse adjustment. However, this approach involves discarding the useful responses that the partial non-respondents did provide. The other is imputation. This approach will simply keep the partial non-respondents in the dataset and fill in all their missing responses. Obviously, it is difficult to make up those missing data but not distort the relationship between all the survey variables.

In DEOCS survey, the missing data comes from the total or unit nonresponse. The average nonresponse rate is about 40%. Weighting procedure is necessary.

3. Type of Weights

Weighting methods can be used for making weighting adjustments for nonresponse and for making certain weighted sample distributions or estimates conform to distributions or estimates obtained from other sources. From Kalton & Kasprzky (1986), the first type of adjustments are called “sample weighting adjustments,” which are for unit nonresponses; the second type are called “population weighting adjustments,” which are for noncoverage. In both cases, the adjustments are used to bring the weighted respondent sample data in line with other data. There are typically seven weighting methods as follows:

1. **Cell weighting.** There are two types of cell weighting. The first and standard cell weighting procedure is called scale weighting. Scale weights adjust the sample weights so that the sample

totals conform to the population totals on a cell-by-cell basis. It will make the number of respondents equal to the number of population. Scale weight calculation is:

$$W_i = \frac{P_i}{R_i}$$

The second is called proportional weighting. Proportional weights are computed from scale weights so that the number of respondents is preserved. Proportional weight calculation is:

$$W_i = \frac{Percent\ of\ Population}{Percent\ of\ Respondents} = \frac{P_i/P_{total}}{R_i/R_{total}}$$

The assumption underlying cell weighting adjustments for nonresponse is that the respondents within a cell represent the non-respondents within that cell. This assumption is satisfied if every element of the population in a cell has the same responding rate if sampled (Kalton & Maligalig, 1991).

2. **Raking.** This is also known as iterative proportional fitting, sample-balancing, or raking ratio estimation. It is a method to adjust the sampling weights of the sample data based on known population characteristics. Whereas cell weighting forces the sample joint distribution of the auxiliary variables to conform to the population joint distribution, raking operates only on the marginal distributions of the auxiliary variables. Raking is an iterative proportional fitting procedure: first, the sample row totals are forced to conform to the population row totals; then the sample adjusted column totals are forced to conform to population column totals; then, the row totals are readjusted to conform; and so on, until convergence is reached. Except for some special cases (Ireland and Kullback, 1968), raking algorithm converges fairly rapidly.
3. **Linear weighting.** Linear weighting belongs to generalized raking methods. The difference is that linear weighting uses a different distance function (Deville, Sarndal, & Sautory, 1993). Like raking, it adjusts the weights to make the sample marginal distributions agree with the population marginal distributions. Linear weighting to marginal distributions is a special case of generalized regression

estimation. Linear weighting has the undesirable feature that it can produce negative weights in some situations. However, some method could be used to place constraints on the adjustments that avoid this outcome (Bethlehem & Keller, 1987; Flores-Cervantes & Kalton, 2003).

4. **Generalized Regression (GREG weighting).** The GREG weighting method is derived from the standard regression estimator in survey sampling (Cochran, 2015). Above weighting methods aim to equate joint or marginal distributions with the population distributions. GREG weighting method is to make weighted sample estimates for quantitative variables conform to population parameters. The method involves incorporating the adjustment for the auxiliary variables used in the regression estimator as a modification of the weights (Baker, Fuller, & McLoughlin, 1994; Bethlehem, 1988; Deville; Flores-Cervantes & Kalton, 2003; Fuller, 2002).

5. **Logistic regression weighting.** Application of logistic regression model in developing weights for adjusting nonresponse was back to 40 years ago (Little, 1986; Little & Vartivarian, 2005; Rosenbaum & Rubin, 1983). With auxiliary variables information from the sample and population, a logistic regression model is constructed to predict the probability of responding, then each respondent's weight is equal to the inverse of the respondent's predicted response probability (Fulsom & Iannacchione, 1991; Lepkowski et al., 1989). When the auxiliary variables in the model are categorical and do not have interactions are included, the weights from logistic regression are similar to raking, but the difference is that they are always larger than 1. Furthermore, compared with raking weighting, logistic regression way is more flexible since its predictors can be continuous, and can also include interaction terms if needed to best predict the response probabilities.

6. **Classification trees.** With a dataset of responders and non-responders, based on auxiliary information, which are usually demographic variables, classification tree could be used to generate a rule to classify responders and non-responders, in fact which will give a probability of each individual of being responders or non-responders. Like logistic regression, the inverse of probability

is the weight for this case. In logistic regression we have to hypothesize specific variables at the beginning and assume that these are the only causes of nonresponse; and if the more auxiliary variables are included in the model, then more problems will come, which like interaction terms, missing data, issues of multicollinearity, and interpretation of the results (Phipps & Toth, 2012). However, classification trees approach can include a large number of auxiliary variables, automatically detect significant interaction effects worth exploring, automatically include item missingness as an indicator of nonresponse, and use multicollinearity to our benefit by allowing variables that are highly correlated to work as surrogates when other variables are missing. Thus, classification trees offer a number of advantages over logistic regression (Earp, Kott, Kreuter, Mitchell, & Porter 2012): 1) classification trees automatically detect significant relationships and interaction effects without pre-specification, reducing the risk of selecting the wrong variables or other model specification errors; 2) the classification tree models identify both the variables that are correlated with the target variable, but also the optimal breakpoints within these variables for maximizing their correlation; 3) they identify hierarchical interaction effects across numerous variables and summarize them using a series of simple rules; 4) they incorporate missing data into the model and assess whether missingness on a given variable is related to the target; 5) they create a series of simple rules that are easy to interpret and use for identifying subgroups with higher propensities; and 6) they reduce the subjectivity of selecting variables to include in the model.

7. **Mixture of cell weighting and another method.** Cell weighting has the advantage over other methods, which does not need other assumptions besides the missing at random (MAR), which are common to all the methods under review. However, it may not be usable when there exist none responses in some cells, and it can also produce unstable weighting adjustments, particularly when the sample sizes in some cells are small (Flores-Cervantes & Kalton, 2003). A compromise approach is to use cell weighting for cells with large sample sizes and some other weighting method for other cells.

4. Computing Weights from Decision Tree in SPSS

SPSS cannot create weights, so we need to create weighting variables on our own. However, once we have computed weights for each ease, SPSS can determine cell counts, and SPSS can apply weights for all analyses. In the dataset of this project, there are four weights as follows:

NonRespWeightCHAID_SAPR

NonRespWeightCHAID_DSPO

NonRespWeightCHAID_ODMEO

NonRespWeightCHAID_TrueIndicator

The weights are created by decision tree (CHAID). Following is the process to organize dataset and create decision tree in this project.

Step 1: create variable “CHAIDGroup” (Graph 1)

Step 2: apply Decision Tree module to generate response probability for each case. The decision tree has been applied to each dataset in Table 1, each of which is for one Service.

Step 3: find weights by computing the reciprocal of response probability.

Table 1. Dataset and Variables for Decision Tree

Dataset	Independent Variables	Dependent Variables
JOINT	Group.Weight; gender; UnitTypeByService	indicator_SAPR
	Group.Weight; gender; UnitTypeByService	indicator_DSPO
	Group.Weight; gender; UnitTypeByService	indicator_ODMEO
	Group.Weight; gender; UnitTypeByService	indicator
NAVY	Group.Weight; gender; CHAIDGROUP	indicator_SAPR
	Group.Weight; gender; CHAIDGROUP	indicator_DSPO
	Group.Weight; gender; CHAIDGROUP	indicator_ODMEO
	Group.Weight; gender; CHAIDGROUP	indicator
USAF	Group.Weight; gender; CHAIDGROUP	indicator_SAPR
	Group.Weight; gender; CHAIDGROUP	indicator_DSPO
	Group.Weight; gender; CHAIDGROUP	indicator_ODMEO
	Group.Weight; gender; CHAIDGROUP	indicator
USA	Group.Weight; gender; CHAIDGROUP	indicator_SAPR
	Group.Weight; gender; CHAIDGROUP	indicator_DSPO
	Group.Weight; gender; CHAIDGROUP	indicator_ODMEO
	Group.Weight; gender; CHAIDGROUP	Indicator
USMC	Group.Weight; gender; CHAIDGROUP	indicator_SAPR

	Group.Weight; gender; CHAIDGROUP	indicator_DSPO
	Group.Weight; gender; CHAIDGROUP	indicator_ODMEO
	Group.Weight; gender; CHAIDGROUP	indicator
USMC	Group.Weight; gender; CHAIDGROUP	indicator_SAPR
	Group.Weight; gender; CHAIDGROUP	indicator_DSPO
	Group.Weight; gender; CHAIDGROUP	indicator_ODMEO
	Group.Weight; gender; CHAIDGROUP	indicator

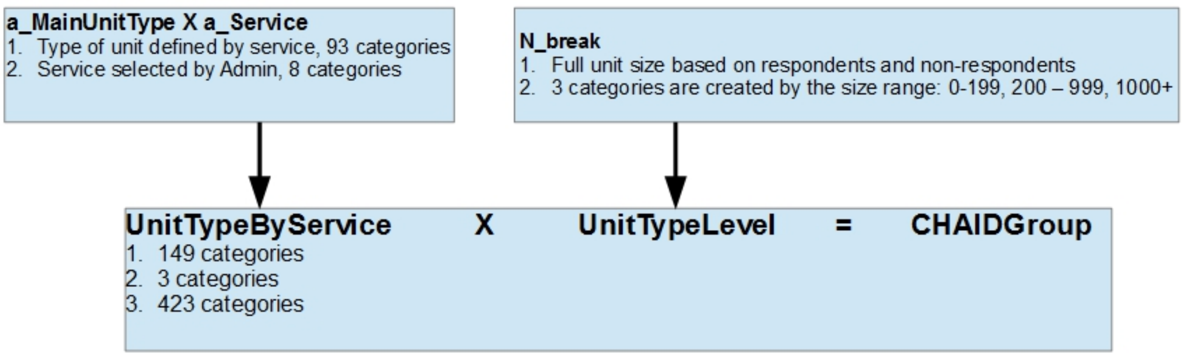


Figure 1. Computing Variable "CHAIDGroup"

5. Nonresponse Bias

The main purpose of weighting adjustments is to reduce the bias in the survey estimates that nonresponse and noncoverage can cause. Thus, we can relate the effects of weighting positively to the nonresponse bias in the survey. However, making the adjustments will lead to increased variability in the weights and thereby lower the precision of the survey estimates. In the case of nonresponse adjustments, a useful measure of this loss of precision is $F = 1 + CV(w_i)^2$, where $CV(w_i)$ is the coefficient of variation of the weights w_i (Kish, 1992). The measure F represents the multiplying factor that is applied to the variance of a survey estimate due to the variability in the weights in the situation where equal weights are optimal (Flores-Cervantes & Kalton, 2003). Office of People Analytics (OPA) uses following four ways to evaluate NRB:

(1) comparing the composition of the sample compared with survey respondents by key demographics

(2) comparing weighted survey estimates of sexual assaults to actual reports

(3) comparing estimates from the NRB follow-up survey (2016 WGRA-N6) to the 2016 WGRA

(4) evaluating the sensitivity of different post-survey adjustments (weighting methods) on survey estimates

In above four methods, Method (1) and (2) are appropriate for this project. Furthermore, we examine the absolute relative bias statistic (hereafter absolute relbias), or the absolute difference between the respondent estimate and the full sample estimate divided by the full sample estimate:

$$R_i = \frac{|\theta_{ri} - \theta_{ni}|}{\theta_{ni}}$$

in which R_i is the absolute relbias for statistic i, θ_{ri} is the estimated value for that statistic based on the respondents, and θ_{ni} is the full sample estimate.

At this stage, we evaluate nonresponse bias from three aspects:

(1) Compare the distribution of variables graphically before and after weighting

(2) Compare the distribution of variables numerically before and after weighting

(3) Compute the absolute relbias statistic for the mean of the above key survey variables and key demographics (Branch, a_Service, Group.Weight, Gender, Race).

5.1 Nonresponse Bias Under Un-Normalized Weights

Since the weights from decision tree are always greater than 1, the weighted total will be greater than the original total. It is a general practice in literature to adjust this difference by multiplying the weights by a constant

$$c = \frac{\text{unweighted N}}{\text{weighted N}}.$$

This process is called normalizing weights. This will bring the N back to the actual (unweighted) N. It does not have any effects on the standard errors if they are calculated using robust standard errors as commonly used in the statistical packages such as Stata and SAS. In the following, we will investigate the nonresponse bias for unweighted and weighted total, separately.

The design of following Algorithm 1 is as follows:

(1) separate the original datafile by "Indicator_SAPR, Indicator_ODMEO, Indicator_DSPO";

(2) compute distribution and other interested statistics for each required variable;

(3) compute distribution and statistics for each variable after weighting;

(4) retrieve the key statistics before and after weighting, compute the absolute relbias and output the results.

Above design is used to develop Algorithm 1 as follows.

5.1.1 Algorithm 1: Weighting, and Evaluating Nonresponse Bias for unweighted total

Input:

Original dataset: "01. Cumulative with NR weights October 2017 - March 2018_For Dr. Wan.sav"

Weighting variables:

Commitment_Fav,SrLeadership_Fav,OrgPerf_Fav,GrpCoh_Fav,Trustlead_Fav,JobSat_Fav,OrgProc_Fav,Eng_Fav, Inclusion_Fav, SexHar_Fav, Prevent_Fav, Response_Fav, SARetaliat_Fav, SHRetaliat_Fav

Output:

Weighting results for above variables and Nonresponse Bias Analysis Results

Method:

Step 1: Create separate datasets according to indictors: Indicator_SAPR, Indicator_ODMEO, Indicator_DSPO;

Step 2: For each dataset from step 1, iterate in weighting variables list to perform following analysis:

Before weighting, run following commands:

(1) FREQUENCIES

(2) DESCRIPTIVES

(3) GGRAPH

Get values of key statistics (mean, variance) of this variable from XML workspace

Apply weights by WEIGHT command, and fun following commands:

(1) FREQUENCIES

(2) DESCRIPTIVES

(3) GGRAPH

Get values of key statistics (mean, variance) of this variable from XML workspace

Create comparison table by CTABLES command to compare effects of weighting.

Compute the absolute relbias for statistics.

Step 3: Output and save dataset file with these variables: weight_variables, mean_beforeweight, mean_afterweight, bias_ratio.

5.1.2 Numerical Results and Conclusion

The graphical and numerical comparison results for each variable are saved in separate excel files (DSPO.xlsx, ODMEO.xlsx, SAPR.xlsx). Direct comparing the graph and numerical distribution of these variables, the effects of weighting are very little.

Following tables are the absolute relbias for statistic for each report (DSPO, ODMEO, SAPR). They are saved as three separate files:

dataset1_indicatorSAPRequals1_BiasRatio.sav,

dataset1_indicatorODMEOequals1_BiasRatio.sav,

dataset1_indicatorDSPOequals1_BiasRatio.sav

From the last column of following Table 2 3 4, the absolute relbias R_i ranges from 0.5% to 1.2%, less than 5%. Thus, the weighting effects are minor. We can deduce that the bias in the survey is little.

Table 2. Nonresponse Bias Ratio for SAPR with Non-normalized Weights

weight_variables	mean_beforeweight	mean_afterweight	bias_ratio
Commitment	71.15	70.50	**0.9290%**
SrLeadersh	73.09	72.22	**1.2033%**
OrgPerf_Fa	71.98	71.30	**0.9566%**
GrpCoh_Fav	71.91	71.29	**0.8712%**
Trustlead_	71.91	71.29	**0.8712%**
JobSat_Fav	68.54	68.27	**0.3848%**
OrgProc_Fa	66.32	65.30	**1.5673%**
Eng_Fav	77.43	77.26	**0.2179%**
Inclusion_	69.29	68.55	**1.0774%**
SexHar_Fav	78.25	77.69	**0.7156%**
Prevent_Fa	81.67	81.17	**0.6106%**
Response_F	87.58	87.13	**0.5182%**
SARetaliat	78.35	77.75	**0.7627%**
SHRetaliat	80.56	79.98	**0.7255%**

Table 3. Nonresponse Bias Ratio for ODME with Non-normalized Weights

weight_variables	mean_beforeweight	mean_afterweight	bias_ratio
Commitment	70.90	70.28	**0.8855%**
SrLeadersh	72.84	71.99	**1.1865%**
OrgPerf_Fa	71.72	71.07	**0.9256%**
GrpCoh_Fav	71.65	71.06	**0.8286%**
Trustlead_	71.65	71.06	**0.8286%**
JobSat_Fav	68.31	68.09	**0.3224%**
OrgProc_Fa	66.08	65.07	**1.5585%**
Eng_Fav	77.15	77.03	**0.1583%**
Inclusion_	68.98	68.28	**1.0280%**
SexHar_Fav	77.76	77.23	**0.6808%**
Prevent_Fa	81.06	80.58	**0.5957%**
Response_F	86.93	86.50	**0.4964%**
SARetaliat	77.70	77.16	**0.6962%**
SHRetaliat	79.91	79.38	**0.6767%**

Table 4. Nonresponse Bias Ratio for DSPO with Non-normalized Weights

weight_variables	mean_beforeweight	mean_afterweight	bias_ratio
Commitment	69.77	69.23	**0.7741%**
SrLeadersh	71.65	70.87	**1.1004%**
OrgPerf_Fa	70.47	69.91	**0.8129%**
GrpCoh_Fav	70.38	69.89	**0.7066%**
Trustlead_	70.38	69.89	**0.7066%**
JobSat_Fav	67.20	67.09	**0.1563%**
OrgProc_Fa	64.77	63.84	**1.4530%**
Eng_Fav	75.93	75.91	**0.0339%**
Inclusion_	67.75	67.19	**0.8409%**
SexHar_Fav	76.47	76.04	**0.5694%**
Prevent_Fa	79.69	79.32	**0.4748%**
Response_F	85.46	85.14	**0.3770%**
SARetaliat	76.47	76.08	**0.5202%**
SHRetaliat	78.66	78.26	**0.5087%**

5.2 Compute Normalized Weights

The purpose of normalization of weights is to set the weights so the total N in the weighted data equals the N in the unweighted data. The method to calculate is to multiply the weight by

$$c = \frac{\text{unweighted N}}{\text{weighted N}}.$$

5.2.1 Design of numerical method:

The input file is the original dataset file.

The first step is to aggregate with break variable “a_DEOCSID7”, and for each case (i.e. each unit) we compute the difference before and after weighting. The output file is Table 5.

The second step is to aggregate with break variable “a_Service” with the input datafile from the first step. For each case (i.e. each service), we compute the difference of totals between before and after weighting, compute normalization ratio for each report. The output file is Table 6.

The third step is to compute normalized weights. The input of this step is the original dataset file and the dataset file from second step. The output file is Table 7, where we can see the extra variables for normalized weights.

Above design is realized in following Algorithm 2.

5.2.2 Algorithm 2: Compute Normalized Weights

Input:

Original dataset: “01. Cumulative with NR weights October 2017 - March 2018_For Dr. Wan.sav”

Output:

Table 5: The dataset file with variable “a_DEOCSID7” as case ID, and variables for the unit totals before and after weighting, their difference, and service ID.

Table 6: The dataset file with variable “a_Service” as case ID, and variables for the service totals before and after weighting, their difference, and the ratio for normalizing weights.

Table 7: The dataset file contains original dataset and normalized weights.

Method:

Step 1: Read in original dataset file.

Iterate through the variable list to create a list of indexes for variables:

"a_DEOCSID7", "TotalAdminRequested_Replaced", "NonRespWeightCHAID_SAPR"

"NonRespWeightCHAID_DSPO", "NonRespWeightCHAID_ODMEO", "a_Service"

Iterate through all cases in the dataset to create dictionaries for above varables:

{"a_DEOCSID7”: The total number of cases in this unit}

{"a_DEOCSID7”: The sum of NonRespWeightCHAID_SAPR in this unit}

{"a_DEOCSID7”: The sum of NonRespWeightCHAID_DSPO in this unit}

{"a_DEOCSID7”: The sum of NonRespWeightCHAID_ODMEO in this unit}

{"a_DEOCSID7”: The a_Service of this unit belongs to}

Iterate through above dictionaries to compute difference of unit totals before and after weighting, and to create following dictionaries to contain these differences:

{"a_DEOCSID7”: The difference of total after NonRespWeightCHAID_SAPR for this unit}

{"a_DEOCSID7”: The difference of total after NonRespWeightCHAID_DSPO for this unit}

{"a_DEOCSID7”: The difference of total after NonRespWeightCHAID_ODMEO for this unit}

Use above dictionaries to create a new datafile (Table 5)

Step 2: Iterate the generated data file from Step 1 to create following dictionaries:

{“a_Service”: The total number of cases in this service}

{“a_Service”: The difference of total after *NonRespWeightCHAID_SAPR* in this service}

{“a_Service”: The difference of total after *NonRespWeightCHAID_DSPO* in this service}

{“a_Service”: The difference of total after *NonRespWeightCHAID_ODMEO* in this service}

{“a_Service”: The normalized ratio for weight-*NonRespWeightCHAID_SAPR* in this service}

{“a_Service”: The normalized ratio for weight-*NonRespWeightCHAID_DSPO* in this service}

{“a_Service”: The normalized ratio for weight-*NonRespWeightCHAID_ODMEO* in this service}

Use above dictionaries to create a new datafile (Table 6)

Step 3: Iterate the original data set, compute three new variables for normalized weights by the dictionaries created in Step 2. And output new dataset (Table 7).

5.2.3 Numerical Results and Conclusion:

The output data file of Algorithm 2 are as follows:

Table 5. Aggregation on a_DEOCSID7 with Extra Variable

dataset1_DifferenceofTotalWithServiceID_ALL3.sav

	a_DEOCSID7	TotalAdminRequested_Replaced	TotalAfterWeighting_SAPR	TotalAfterWeighting_DSPO	TotalAfterWeighting_ODMEO	TotalAfterMinusBefore_SAPR	TotalAfterMinusBefore_DSPO	TotalAfterMinusBefore_ODMEO	a_Service
1	1802241.00	89.00	80.22	-8.78	75.89	-13.11	80.82	-8.18	3.00
2	1802242.00	724.00	503.99	-220.01	453.43	-270.57	501.92	-222.08	4.00
3	1802243.00	85.00	165.94	80.94	151.44	66.44	163.74	78.74	4.00
4	1802244.00	514.00	803.03	289.03	755.42	241.42	799.41	285.41	4.00
5	1802246.00	76.00	88.68	12.68	85.95	9.95	91.44	15.44	4.00
6	1802251.00	167.00	300.12	133.12	271.33	104.33	297.22	130.22	3.00
7	1802252.00	23.00	47.44	24.44	43.23	20.23	47.21	24.21	4.00
8	1802253.00	745.00	547.15	-197.85	498.31	-246.69	539.85	-205.15	2.00
9	1802254.00	193.00	144.58	-48.42	131.87	-61.13	142.32	-50.68	4.00
10	1802256.00	70.00	43.09	-26.91	39.66	-30.34	42.49	-27.51	4.00
11	1802262.00	184.00	232.43	48.43	209.48	25.48	228.06	44.06	4.00
12	1802264.00	64.00	106.28	42.28	96.97	32.97	104.04	40.04	4.00

The variables and labels for above data set are as follows.

Variables	Label
a_DEOCSID7	Unit ID
TotalAdminRequested_Replaced	Unit total
TotalAfterWeighting_SAPR	Unit total after weighting for SAPR
TotalAfterWeighting_DSPO	Unit total after weighting for DSPO
TotalAfterWeighting_ODMEO	Unit total after weighting for ODMEO
TotalAfterMinusBefore_SAPR	Difference casued by weighting for SAPR
TotalAfterMinusBefore_DSPO	Difference casued by weighting for DSPO
TotalAfterMinusBefore_ODMEO	Difference casued by weighting for ODMEO
a_Service	Service

In above dataset, for each unit (a_DEOCSID7), the algorithm computes the difference between weighted and unweight totals, and also attaches the service (a_Service) which this unit belongs to. The "unusual" thing about this datafile is: *Total After Weighting – Total before Weighting < 0* for some cases. The reason is that the weights are generated for service-wise instead of unit-wise. Thus, we will compute this quantity and normalize weights for each Service.

Table 6. Aggregation on a_Service with Extra Variable

dataset1_DifferenceofTotalbyService_ALL3.sav

	a_Service	TotalBeforeWeighting	TotalAfterWeightingSAPR	AfterMinusBeforeSAPR	AfterBeforeRatioSAPR	NormalizationRatioSAPR	TotalAfterWeightingDSPO	AfterMinusBeforeDSPO	AfterBeforeRatioDSPO	NormalizationRatioDSPO	TotalAfterWeightingODMEO	AfterMinusBeforeODMEO	AfterBeforeRatioODMEO	NormalizationRatioODMEO
1	1.00	37107.41	45102.43	7995.01	.2155	.8227	42010.94	4903.53	.1321	.8833	44985.87	7878.46	.2123	.8249
2	2.00	289495.33	332041.30	42545.98	.1470	.8719	311997.87	22502.54	.0777	.9279	330205.00	40709.68	.1406	.8767
3	3.00	329868.91	394387.12	64518.20	.1956	.8364	364463.94	34595.03	.1049	.9051	392810.88	62941.97	.1908	.8398
4	4.00	528052.48	626379.00	98326.52	.1862	.8430	577840.42	49787.95	.0943	.9138	618245.25	90192.77	.1708	.8541
5	5.00	83561.45	97446.29	13884.85	.1662	.8575	90640.44	7079.00	.0847	.9219	96144.40	12582.96	.1506	.8691
6	8.00	123679.93	142590.46	18910.54	.1529	.8674	133626.51	9946.59	.0804	.9256	141589.57	17909.64	.1448	.8735

The variables and labels for above data set are as follows.

Variables	Label
a_Service	Unit ID
TotalBeforeWeighting	Service total
TotalAfterWeightingSAPR	Service total after weighting SAPR
AfterMinusBeforeSAPR	Difference casued by weighting SAPR
AfterBeforeRatioSAPR	Difference ratio casued by weighting SAPR
NormalizationRatioSAPR	Normalization ration for SAPR
TotalAfterWeightingDSPO	Service total after weighting DSPO
AfterMinusBeforeDSPO	Difference casued by weighting DSPO
AfterBeforeRatioDSPO	Difference ratio casued by weighting DSPO
NormalizationRatioDSPO	Normalization ration for DSPO
TotalAfterWeightingODMEO	Service total after weighting ODMEO
AfterMinusBeforeODMEO	Difference casued by weighting ODMEO
AfterBeforeRatioODMEO	Difference ratio casued by weighting ODMEO
NormalizationRatioODMEO	Normalization ration for ODMEO

In above dataset, for each service, the algorithm computes the difference between weighted and unweight totals, and the ratio for normalizing weights. Then we insert the normalized weights back to the original dataset for further analysis. Following dataset (Table 7) contains all variables and cases from original dataset "01. Cumulative with NR weights October 2017 - March 2018_For Dr. Wan", and normalized weights for SAPR, DSPO, ODMEO.

Table 7. Original Dataset with Normalized Weights

dataset1_AllDatasetwithNormedWeights1.sav

	NonRespWeightCHAID_SAPR	NormedWeightSAPR	NonRespWeightCHAID_DSPO	NormedWeightDSPO	NonRespWeightCHAID_ODMEO	NormedWeightODMEO
1	1.73	1.51	1.66	1.54	1.72	1.51
2	2.15	1.87	2.08	1.93	2.13	1.87
3	1.74	1.52	1.73	1.61	1.74	1.53
4	1.78	1.55	1.72	1.59	1.76	1.54
5	1.73	1.51	1.66	1.54	1.72	1.51
6	1.73	1.51	1.66	1.54	1.72	1.51
7	1.78	1.55	1.72	1.59	1.76	1.54
8	1.73	1.51	1.66	1.54	1.72	1.51
9	2.15	1.87	2.08	1.93	2.13	1.87
10	1.73	1.51	1.66	1.54	1.72	1.51
11	1.73	1.51	1.66	1.54	1.72	1.51
12	2.15	1.87	2.08	1.93	2.13	1.87

5.3 Compute Nonresponse Bias with Normalized Weights

Algorithm 2 has generated a dataset file, which contains the normalized weights for each report. In this part, we will repeat the analysis conducted in Algorithm 1 with normalized weights.

5.3.1 Design of numerical method:

The input file is the output dataset file (dataset1_AllDatasetwithNormedWeights1.sav) from Algorithm 2. The design of Algorithm 3 is similar to Algorithm 1. We need to set up an iteration to finish following comparisons for each report-SAPR, ODMEO, DSPO:

(1) compute distribution and other interested statistics for each required variable;

(2) compute the same distribution and statistics for each variable after weighting;

(3) retrieve the key statistics before and after weighting, compute the absolute relbias and output the results.

Above idea is used to develop Algorithm 3 as follows.

5.3.2 Algorithm 3: Weighting, and Evaluating Nonresponse Bias with Normalized Weights

Input:

Dataset from Algorithm 2: “dataset1_AllDatasetwithNormedWeights1.sav”

Output:

Table 8: a data set with weighted variables and its absolute relbias for SAPR.

Table 9: a data set with weighted variables and its absolute relbias for ODMEO.

Table 10: a data set with weighted variables and its absolute relbias for DSPO.

Method:

Iterate through the indicator list: *["Indicator_SAPR", "Indicator_ODMEO", "Indicator_DSPO"]*

Generate a data set for this indicator

Iterate through the following variable list to run following commands:

["Commitment_Fav", "SrLeadership_Fav","OrgPerf_Fav", "GrpCoh_Fav","Trustlead_Fav", "JobSat_Fav", "OrgProc_Fav","Eng_Fav", "Inclusion_Fav", "SexHar_Fav", "Prevent_Fav", "Response_Fav", "SARetaliat_Fav","SHRetaliat_Fav", "Branch","a_Service", "Group.Weight", "Gender", "Race"]

FREQUENCIES

GGRAPH

DESCRIPTIVES VARIABLES

WEIGHT BY (variable name).

FREQUENCIES

GGRAPH

DESCRIPTIVES VARIABLES

CTABLES (create custom table to compute absolute relbias)

WEIGHT OFF.

Create and output a data set with weighted variables and its absolute relbias for each indicator.

5.3.3 Numerical Results and Conclusion:

The output data file of Algorithm 3 are as follows:

Table 8. Nonresponse Bias Ratio for SAPR with Normalized Weights

weight_variables	mean_beforeweight	mean_afterweight	bias_ratio
Commitment	71.15	70.51	**0.909%**
SrLeadersh	73.09	72.24	**1.175%**
OrgPerf_Fa	71.98	71.32	**0.927%**
GrpCoh_Fav	71.91	71.30	**0.853%**
Trustlead_	71.91	71.30	**0.853%**
JobSat_Fav	68.54	68.29	**0.361%**
OrgProc_Fa	66.32	65.32	**1.536%**
Eng_Fav	77.43	77.26	**0.216%**
Inclusion_	69.29	68.56	**1.067%**
SexHar_Fav	78.25	77.69	**0.717%**
Prevent_Fa	81.67	81.17	**0.618%**
Response_F	87.58	87.13	**0.517%**
SARetaliat	78.35	77.75	**0.770%**
SHRetaliat	80.56	79.98	**0.731%**
Branch	80.56	79.98	**0.731%**
a_Service	80.56	79.98	**0.731%**
Group.Weig	80.56	79.98	**0.731%**
Gender	1.21	1.22	**0.714%**
Race	1.21	1.22	**0.714%**

Table 9. Nonresponse Bias Ratio for DSPO with Normalized Weights

weight_variables	mean_beforeweight	mean_afterweight	bias_ratio
Commitment	69.77	69.24	**0.772%**
SrLeadersh	71.65	70.88	**1.092%**
OrgPerf_Fa	70.47	69.91	**0.808%**
GrpCoh_Fav	70.38	69.89	**0.705%**
Trustlead_	70.38	69.89	**0.705%**
JobSat_Fav	67.20	67.09	**0.153%**
OrgProc_Fa	64.77	63.85	**1.436%**
Eng_Fav	75.93	75.90	**0.044%**
Inclusion_	67.75	67.19	**0.845%**
SexHar_Fav	76.47	76.03	**0.576%**
Prevent_Fa	79.69	79.31	**0.485%**
Response_F	85.46	85.13	**0.384%**
SARetaliat	76.47	76.06	**0.537%**
SHRetaliat	78.66	78.25	**0.522%**
Branch	78.66	78.25	**0.522%**
a_Service	78.66	78.25	**0.522%**
Group.Weig	78.66	78.25	**0.522%**
Gender	1.21	1.22	**0.595%**
Race	1.21	1.22	**0.595%**

Table 10. Nonresponse Bias Ratio for ODMEO with Normalized Weights

weight_variables	mean_beforeweight	mean_afterweight	bias_ratio
Commitment	70.90	70.29	**0.876%**
SrLeadersh	72.84	72.00	**1.171%**
OrgPerf_Fa	71.72	71.08	**0.911%**
GrpCoh_Fav	71.65	71.06	**0.821%**
Trustlead_	71.65	71.06	**0.821%**
JobSat_Fav	68.31	68.09	**0.311%**
OrgProc_Fa	66.08	65.08	**1.531%**
Eng_Fav	77.15	77.02	**0.169%**
Inclusion_	68.98	68.27	**1.030%**
SexHar_Fav	77.76	77.22	**0.691%**
Prevent_Fa	81.06	80.57	**0.611%**
Response_F	86.93	86.49	**0.505%**
SARetaliat	77.70	77.14	**0.720%**
SHRetaliat	79.91	79.36	**0.696%**
Branch	79.91	79.36	**0.696%**
a_Service	79.91	79.36	**0.696%**
Group.Weig	79.91	79.36	**0.696%**
Gender	1.21	1.22	**0.707%**
Race	1.21	1.22	**0.707%**

5.3.3 Numerical Results and Conclusion:

The observations are as follows:

(1) Direct comparing the graph and numerical distribution of these weighted variables, the effects of weighting are little. From the last column of following tables, the absolute relbias R_i ranges from 0.3% to 1.5%. Especially, the key demographical variables (Branch, a_Service, Group.Weight, Gender, Race)'s R_i ranges from 0.5% to 0.7%. Thus, the weighting effects are minor, and the nonresponse bias is little.

(2) The difference of absolute relbias R_i between original weights and normalized weights is very little.

6. Evaluation of Weights by Trendline

One way to evaluate weights is to check the trend line of the key survey weighted estimates overall and in key subpopulations to see whether the changes across waves are reasonable. In the dataset of this project, it contains a variable “month”, which has following basic distribution (Table 11). Thus, the original dataset contains the survey from six time periods, which enable us to check the change of weighting effects over these time periods. Besides evaluating the weight itself, this comparison will also be used as a method to check non-response bias. The major indexes for evaluating weights are the mean and standard deviation of those key survey variables.

Table 11. Distribution of Variable "Month"

Month		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	October 2017	124748	20.4	20.4	20.4
	November 2017	122209	19.9	19.9	40.3
	December 2017	109236	17.8	17.8	58.1
	Jan-18	82799	13.5	13.5	71.6
	February 2018	83706	13.7	13.7	85.3
	March 2018	90123	14.7	14.7	100.0
	Total	612821	100.0	100.0	

6.1 Design of numerical method:

The input file is the original dataset file.

The first iteration is in the list of *["Indicator_SAPR", "Indicator_ODMEO", "Indicator_DSPO"]*, which means we will evaluate the trend line for each report.

Generate a separate dataset for each indicator.

Set weight on

Start second iteration in the list of all variables:

["Commitment_Fav","SrLeadership_Fav","OrgPerf_Fav", "GrpCoh_Fav","Trustlead_Fav", "JobSat_Fav", "OrgProc_Fav","Eng_Fav", "Inclusion_Fav", "SexHar_Fav", "Prevent_Fav", "Response_Fav", "SARetaliat_Fav","SHRetaliat_Fav"]

Compute the key statistics of these variables. Output the results as graph and save them as a dictionary.

In each report, compute the mean and standard deviation of each survey variable and output them as a dataset file.

6.2 Algorithm 4: Weighting Evaluation by Checking Trend Line

Input:

Original Dataset from Algorithm 2: “dataset1_AllDatasetwithNormedWeights1.sav”

Output:

Table 12: Mean and standard deviation of key survey variables over time period for SAPR

Table 13: Mean and standard deviation of key survey variables over time period for DSPO

Table 14: Mean and standard deviation of key survey variables over time period for ODMEO

Method:

Iterate through the indicator list: *["Indicator_SAPR", "Indicator_ODMEO", "Indicator_DSPO"]*

Generate a separate dataset for each indicator.

Set weight on.

Start second iteration in the list of all key survey variables:

["Commitment_Fav",…,"SHRetaliat_Fav"]

Compute mean and standard deviation for each category of “*Month*”

Save as lists, and plot the graphs: “mean vs month”, and “standard deviation vs month”

Create two data sets for each indicator: one for mean value, the other for standard deviation. Each case is for each survey variable, and each variable is for each time period.

In data sets generated above, compute z-score for each variable.

Output and save above data set files.

6.3 Numerical Results and Conclusion

Table 13-18 give the mean, standard deviation, and their z-scores of weighted key survey variables over past six periods of time.

Figure 2-4 are plotted by the data from Table 13-18. The x-axis is “Month”, and y-axis are mean or standard deviation for each variable.

The observations (Table 12, Table 13-18) are follows:

1. The mean values of all survey variables fall in 2σ of average of all mean values.

2. The standard deviations of all survey variables fall in 2σ of average of all standard deviation.

Thus, there are no extreme weights.

Table 12: Range of z-score of Mean and Std. Deviation

	Range of z-score of mean	Range of z-score of std. deviation
SAPR	[-1.44, 1.61]	[-1.89, 1.59]
DSPO	[-1.56, 1.88]	[-1.86, 1.53]
ODMEO	[-1.76, 1.69]	[-1.89, 1.56]

Table 13: Mean of key survey variables over time period for SAPR

VariableList	Oct2017_zscore	Nov2017_zscore	Dec2017_zscore	Jan2018_zscore	Feb2018_zscore	Mar2018_zscore	Oct2017	Nov2017	Dec2017	Jan2018	Feb2018	Mar2018	Mon Mean	MonSD
GrpCoh_F	-0.91	1.31	-1.15	0.80	0.48	-0.53	70.86	71.98	70.74	71.72	71.56	71.05	71.32	0.50
OrgPerf_F	0.73	1.48	-0.66	-0.04	-0.18	-1.35	71.73	72.24	70.78	71.20	71.11	70.31	71.23	0.69
Commitm	-1.23	1.45	-0.66	0.90	-0.06	-0.39	69.96	71.20	70.22	70.94	70.50	70.34	70.53	0.46
Eng_Fav	-0.05	1.66	-0.40	0.62	-0.76	-1.06	77.19	77.99	77.02	77.50	76.86	76.71	77.21	0.47
SHRetaliat	0.44	1.43	0.44	-0.26	-0.59	-1.46	80.21	80.97	80.20	79.66	79.40	78.72	79.86	0.78
SexHar_Fa	0.34	1.14	0.83	-0.73	-0.02	-1.54	77.78	78.15	78.00	77.29	77.62	76.93	77.63	0.46
Trustlead_	-1.38	1.61	0.06	0.41	-0.12	-0.58	78.12	79.43	78.75	78.90	78.67	78.47	78.72	0.44
Response_	0.64	1.24	0.66	-0.44	-0.74	-1.37	87.43	87.79	87.44	86.76	86.57	86.19	87.03	0.62
JobSat_Fa	-0.12	1.30	-1.75	0.34	0.32	-0.09	68.22	68.99	67.33	68.47	68.45	68.23	68.28	0.54
SrLeaders	-0.73	0.76	0.17	-1.58	1.13	0.25	72.04	72.44	72.28	71.81	72.54	72.30	72.24	0.27
OrgProc_F	-1.44	0.07	-0.29	0.37	1.61	-0.31	64.63	65.43	65.24	65.59	66.24	65.23	65.39	0.53
SARetaliat	0.60	1.35	0.50	-0.36	-0.68	-1.41	78.14	78.80	78.05	77.28	77.00	76.36	77.60	0.89
Prevent_F	-1.19	0.94	1.32	0.21	-0.87	-0.41	80.73	81.50	81.64	81.24	80.84	81.01	81.16	0.37
Inclusion_	-1.01	1.74	-0.64	0.33	0.20	-0.62	68.19	69.18	68.33	68.68	68.63	68.33	68.56	0.36

Table 14: Standard deviation of key survey variables over time period for SAPR

VariableList	Oct2017_zscore	Nov2017_zscore	Dec2017_zscore	Jan2018_zscore	Feb2018_zscore	Mar2018_zscore	Oct2017	Nov2017	Dec2017	Jan2018	Feb2018	Mar2018	MonMean	MonSD
GrpCoh_F	0.10	-1.83	0.70	-0.04	0.00	1.07	39.25	38.76	39.41	39.22	39.23	39.50	39.23	0.26
OrgPerf_F	-1.02	-1.41	0.32	0.21	0.81	1.09	38.07	37.87	38.73	38.68	38.98	39.11	38.57	0.50
Commitm	0.41	-1.88	0.37	-0.24	0.36	0.98	41.30	40.77	41.29	41.15	41.28	41.42	41.20	0.23
Eng_Fav	-0.46	-1.64	-0.01	0.07	1.07	0.98	32.50	32.03	32.68	32.72	33.12	33.08	32.69	0.40
SHRetaliat	-0.45	-1.48	-0.38	0.30	0.57	1.43	37.21	36.65	37.25	37.61	37.76	38.23	37.45	0.54
SexHar_Fa	-0.57	-1.44	-0.48	0.67	0.53	1.28	31.01	30.66	31.04	31.51	31.45	31.75	31.24	0.40
Trustlead_	0.57	-1.89	-0.02	-0.09	0.51	0.92	35.23	34.54	35.06	35.04	35.21	35.32	35.07	0.28
Response_	-0.83	-1.24	-0.50	0.59	0.67	1.31	27.89	27.57	28.14	28.98	29.04	29.53	28.53	0.77
JobSat_Fa	-0.32	-1.55	1.51	0.29	-0.23	0.30	43.98	43.70	44.41	44.13	44.01	44.13	44.06	0.23
SrLeaders	-0.79	-1.03	-0.60	1.48	0.06	0.88	38.84	38.80	38.88	39.31	39.02	39.19	39.01	0.21
OrgProc_F	-0.91	-0.96	0.77	-0.16	-0.32	1.59	39.30	39.30	39.61	39.44	39.41	39.76	39.47	0.18
SARetaliat	-0.63	-1.41	-0.44	0.61	0.56	1.31	36.76	36.28	36.88	37.52	37.49	37.95	37.15	0.61
Prevent_F	0.02	-1.34	-0.96	0.17	0.96	1.15	29.55	29.17	29.27	29.59	29.81	29.86	29.54	0.28
Inclusion_	-0.50	-1.71	0.15	0.34	0.52	1.19	34.17	33.85	34.35	34.40	34.45	34.63	34.31	0.27

Table 15: Mean of key survey variables over time period for DSPO

VariableList	Oct2017_zscore	Nov2017_zscore	Dec2017_zscore	Jan2018_zscore	Feb2018_zscore	Mar2018_zscore	Oct2017	Nov2017	Dec2017	Jan2018	Feb2018	Mar2018	MonMean	MonSD
GrpCoh_F	-0.52	1.66	-0.92	0.62	0.02	-0.86	69.59	70.76	69.38	70.21	69.88	69.41	69.87	0.54
OrgPerf_F	0.84	1.47	-0.41	-0.08	-0.51	-1.30	70.48	70.99	69.44	69.72	69.36	68.71	69.78	0.82
Commitm	-0.83	1.73	-0.39	0.66	-0.40	-0.77	68.81	70.07	69.03	69.55	69.02	68.84	69.22	0.49
Eng_Fav	0.19	1.63	-0.20	0.37	-0.74	-1.25	75.94	76.78	75.72	76.05	75.40	75.11	75.83	0.58
SHRetaliat	0.63	1.35	0.38	-0.26	-0.64	-1.47	78.69	79.39	78.46	77.84	77.48	76.68	78.09	0.96
SexHar_Fa	0.66	1.15	0.62	-0.59	-0.27	-1.56	76.34	76.64	76.32	75.57	75.77	74.98	75.94	0.61
Trustlead_	-0.75	1.78	0.12	0.31	-0.55	-0.91	76.78	78.14	77.25	77.35	76.89	76.70	77.19	0.54
Response_	0.81	1.24	0.45	-0.37	-0.74	-1.39	85.66	86.02	85.36	84.68	84.37	83.82	84.99	0.83
JobSat_Fa	0.19	1.55	-1.56	0.13	0.00	-0.32	67.17	67.98	66.13	67.13	67.06	66.87	67.06	0.59
SrLeadersl	0.04	1.42	0.40	-1.66	0.07	-0.27	70.85	71.27	70.96	70.33	70.86	70.75	70.84	0.31
OrgProc_F	-1.26	0.74	-0.24	0.30	1.38	-0.91	63.34	64.20	63.78	64.01	64.47	63.49	63.88	0.43
SARetaliat	0.75	1.28	0.42	-0.31	-0.71	-1.43	76.68	77.25	76.33	75.55	75.12	74.34	75.88	1.07
Prevent_F	-0.39	1.29	1.08	0.02	-1.01	-1.00	79.09	79.85	79.75	79.27	78.80	78.81	79.26	0.46
Inclusion_	-0.29	1.88	-0.51	0.17	-0.24	-1.01	67.02	67.96	66.93	67.22	67.04	66.72	67.15	0.43

Table 16: Standard deviation of key survey variables over time period for DSPO

VariableList	Oct2017_zscore	Nov2017_zscore	Dec2017_zscore	Jan2018_zscore	Feb2018_zscore	Mar2018_zscore	Oct2017	Nov2017	Dec2017	Jan2018	Feb2018	Mar2018	MonMean	MonSD
GrpCoh_F	-0.22	-1.81	0.43	0.11	0.31	1.18	39.94	39.46	40.14	40.04	40.10	40.36	40.01	0.30
OrgPerf_F	-1.03	-1.35	0.14	0.23	0.94	1.07	38.79	38.61	39.42	39.47	39.86	39.93	39.35	0.55
Commitm	0.06	-1.86	0.10	0.01	0.64	1.05	41.82	41.32	41.83	41.81	41.98	42.08	41.81	0.26
Eng_Fav	-0.57	-1.57	-0.12	0.20	0.92	1.14	33.57	33.09	33.79	33.94	34.29	34.40	33.85	0.48
SHRetaliat	-0.66	-1.39	-0.31	0.30	0.63	1.44	38.46	38.00	38.68	39.06	39.26	39.77	38.87	0.63
SexHar_Fa	-0.83	-1.31	-0.35	0.48	0.68	1.33	32.46	32.21	32.72	33.16	33.26	33.60	32.90	0.53
Trustlead_	0.02	-1.79	-0.09	0.00	0.79	1.08	36.27	35.59	36.22	36.26	36.56	36.67	36.26	0.38
Response_	-0.95	-1.24	-0.29	0.47	0.67	1.34	30.27	30.00	30.89	31.59	31.78	32.40	31.16	0.93
JobSat_Fa	-0.54	-1.65	1.26	0.39	0.04	0.50	44.39	44.12	44.84	44.63	44.54	44.65	44.53	0.25
SrLeadersl	-0.85	-1.19	-0.58	1.20	0.50	0.92	39.52	39.43	39.60	40.09	39.90	40.01	39.76	0.28
OrgProc_F	-1.06	-1.12	0.43	-0.06	0.29	1.53	39.84	39.82	40.20	40.08	40.16	40.46	40.09	0.24
SARetaliat	-0.80	-1.34	-0.34	0.51	0.61	1.35	37.96	37.59	38.28	38.86	38.93	39.44	38.51	0.69
Prevent_F	-0.64	-1.33	-0.49	0.26	0.87	1.32	31.42	31.11	31.49	31.82	32.10	32.30	31.70	0.45
Inclusion_	-0.80	-1.52	0.10	0.37	0.64	1.22	34.99	34.74	35.31	35.41	35.50	35.71	35.28	0.36

Table 17: Mean of key survey variables over time period for ODMEO

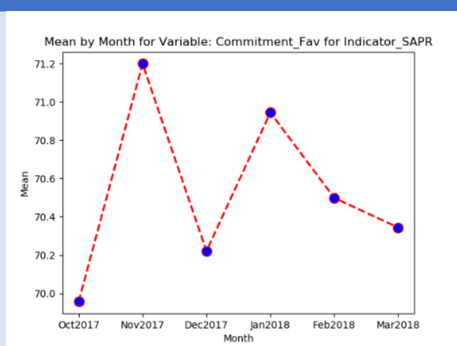
VariableList	Oct2017_zscore	Nov2017_zscore	Dec2017_zscore	Jan2018_zscore	Feb2018_zscore	Mar2018_zscore	Oct2017	Nov2017	Dec2017	Jan2018	Feb2018	Mar2018	MonMean	MonSD
GrpCoh_F	-0.88	1.40	-1.13	0.73	0.42	-0.54	70.64	71.78	70.51	71.44	71.29	70.81	71.08	0.50
OrgPerf_F	0.74	1.50	-0.57	-0.08	-0.23	-1.35	71.49	72.03	70.57	70.92	70.82	70.03	70.98	0.70
Commitm	-1.21	1.52	-0.61	0.82	-0.09	-0.43	69.74	71.00	70.01	70.67	70.25	70.10	70.30	0.46
Eng_Fav	0.04	1.69	-0.41	0.52	-0.79	-1.04	77.00	77.76	76.78	77.22	76.61	76.49	76.98	0.46
SHRetaliat	0.50	1.42	0.44	-0.30	-0.63	-1.43	79.65	80.41	79.60	78.99	78.71	78.04	79.23	0.83
SexHar_Fa	0.44	1.20	0.75	-0.73	-0.16	-1.50	77.37	77.75	77.53	76.79	77.07	76.41	77.15	0.50
Trustlead_	-1.27	1.67	0.11	0.32	-0.17	-0.66	77.86	79.19	78.48	78.57	78.36	78.13	78.43	0.45
Response_	0.68	1.27	0.60	-0.42	-0.78	-1.34	86.84	87.25	86.79	86.09	85.84	85.46	86.38	0.69
JobSat_Fa	-0.06	1.32	-1.76	0.22	0.32	-0.05	68.05	68.78	67.15	68.20	68.25	68.06	68.08	0.53
SrLeadersl	-0.70	0.90	0.25	-1.60	1.04	0.12	71.81	72.22	72.06	71.58	72.26	72.02	71.99	0.26
OrgProc_F	-1.45	0.15	-0.29	0.33	1.60	-0.34	64.41	65.23	65.00	65.32	65.98	64.98	65.15	0.51
SARetaliat	0.63	1.35	0.47	-0.36	-0.72	-1.38	77.59	78.27	77.43	76.64	76.30	75.67	76.98	0.95
Prevent_F	-0.94	1.17	1.21	0.10	-0.97	-0.57	80.18	81.01	81.03	80.59	80.17	80.33	80.55	0.39
Inclusion_	-0.90	1.82	-0.59	0.23	0.11	-0.68	67.94	68.94	68.05	68.35	68.31	68.01	68.27	0.37

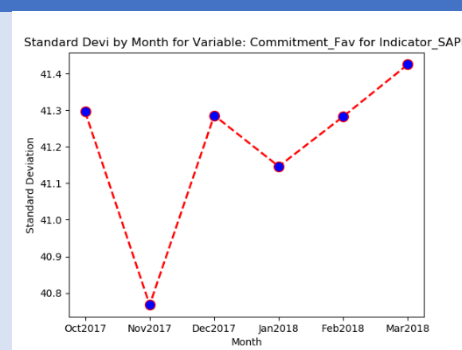
Table 18: Standard deviation of key survey variables over time period for ODMEO

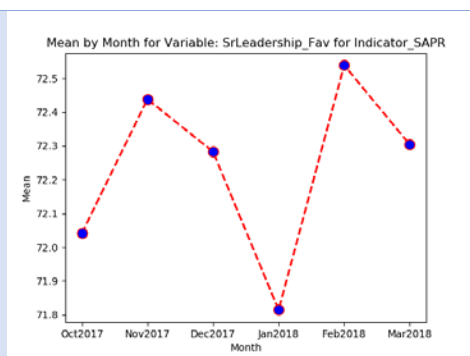
VariableList	Oct2017_zscore	Nov2017_zscore	Dec2017_zscore	Jan2018_zscore	Feb2018_zscore	Mar2018_zscore	Oct2017	Nov2017	Dec2017	Jan2018	Feb2018	Mar2018	MonMean	MonSD
GrpCoh_F	0.05	-1.88	0.65	0.05	0.12	1.01	39.39	38.88	39.54	39.39	39.40	39.64	39.37	0.26
OrgPerf_F	-1.00	-1.41	0.25	0.23	0.86	1.08	38.22	38.01	38.85	38.83	39.15	39.26	38.72	0.50
Commitm	0.39	-1.90	0.31	-0.17	0.42	0.96	41.40	40.88	41.39	41.28	41.41	41.54	41.32	0.23
Eng_Fav	-0.54	-1.62	0.01	0.13	1.07	0.96	32.67	32.22	32.90	32.95	33.33	33.29	32.89	0.41
SHRetaliat	-0.51	-1.46	-0.38	0.34	0.62	1.40	37.68	37.13	37.75	38.16	38.32	38.76	37.97	0.57
SexHar_Fa	-0.66	-1.43	-0.41	0.65	0.62	1.24	31.37	31.04	31.48	31.94	31.93	32.19	31.66	0.43
Trustlead_	0.41	-1.89	-0.07	0.03	0.55	0.97	35.43	34.74	35.29	35.32	35.47	35.60	35.31	0.30
Response_	-0.85	-1.27	-0.44	0.56	0.71	1.29	28.71	28.35	29.04	29.87	29.99	30.47	29.41	0.83
JobSat_Fa	-0.38	-1.55	1.47	0.39	-0.23	0.29	44.05	43.78	44.48	44.23	44.09	44.21	44.14	0.23
SrLeaders	-0.81	-1.06	-0.62	1.38	0.14	0.96	38.98	38.93	39.02	39.44	39.18	39.35	39.15	0.21
OrgProc_F	-0.93	-1.03	0.75	-0.13	-0.22	1.56	39.41	39.39	39.73	39.56	39.54	39.88	39.59	0.19
SARetaliat	-0.67	-1.41	-0.41	0.59	0.60	1.29	37.22	36.73	37.38	38.03	38.03	38.48	37.65	0.64
Prevent_F	-0.24	-1.44	-0.72	0.27	0.96	1.16	30.19	29.77	30.02	30.37	30.61	30.68	30.27	0.35
Inclusion_	-0.56	-1.68	0.09	0.43	0.56	1.16	34.31	33.99	34.49	34.59	34.62	34.79	34.46	0.28

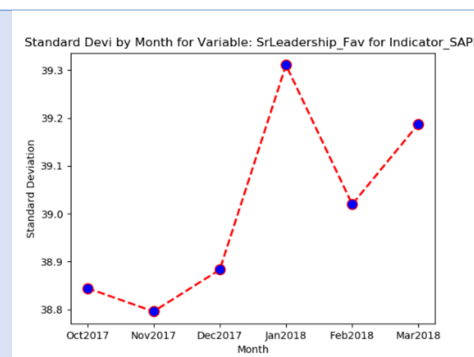
Figure 2: Mean and standard deviation of key survey variables over time period for SAPR

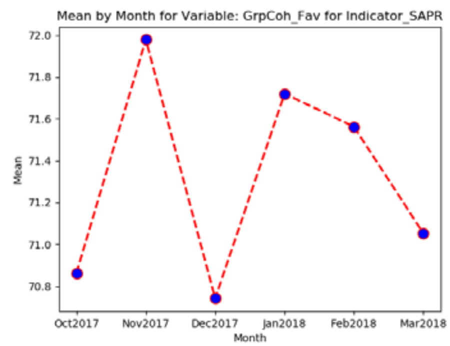
Mean Value for Survey Variable	Standard Deviation for Survey Variable

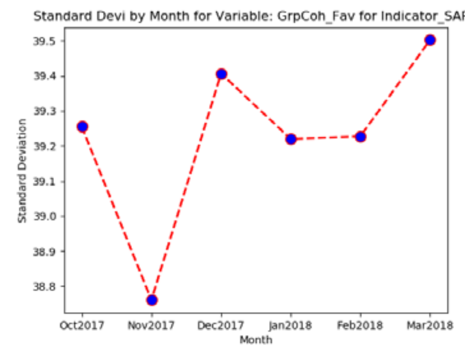


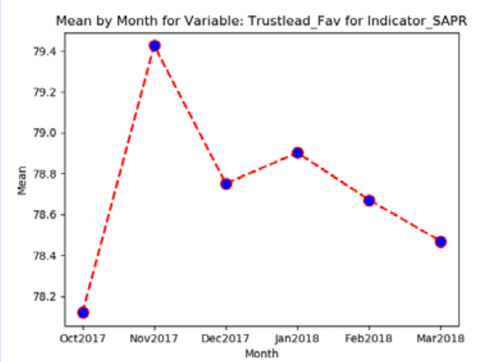


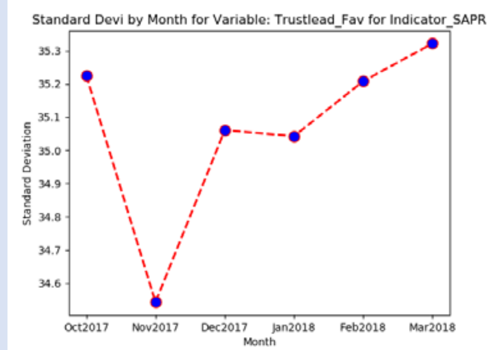


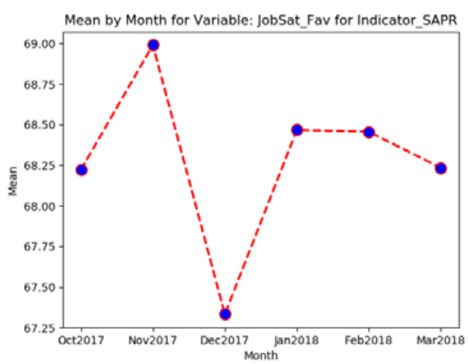


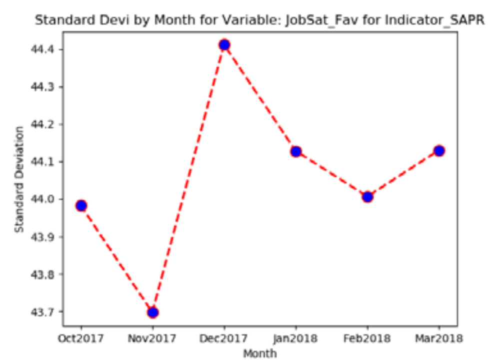

Mean by Month for Variable: GrpCoh_Fav for Indicator_SAPR
Mean
72.0
71.8
71.6
71.4
71.2
71.0
70.8
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

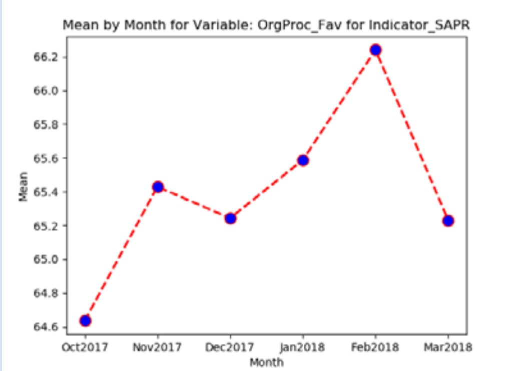

Standard Devi by Month for Variable: GrpCoh_Fav for Indicator_SAPR
Standard Deviation
39.5
39.4
39.3
39.2
39.1
39.0
38.9
38.8
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

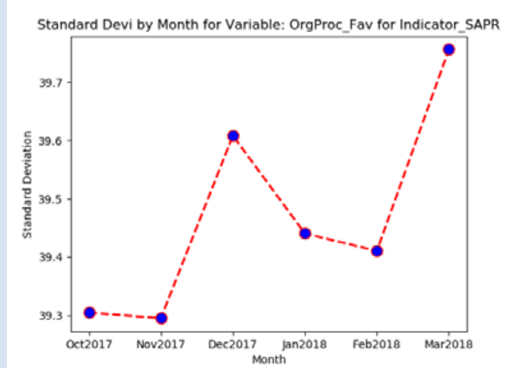

Mean by Month for Variable: Trustlead_Fav for Indicator_SAPR
Mean
79.4
79.2
79.0
78.8
78.6
78.4
78.2
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

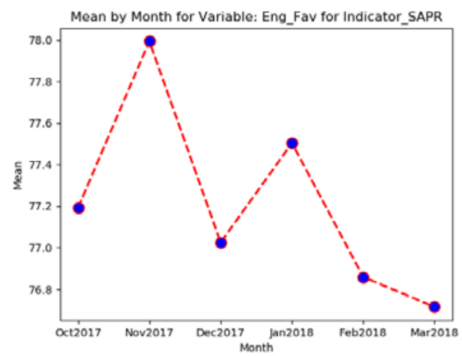

Standard Devi by Month for Variable: Trustlead_Fav for Indicator_SAPR
Standard Deviation
35.3
35.2
35.1
35.0
34.9
34.8
34.7
34.6
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

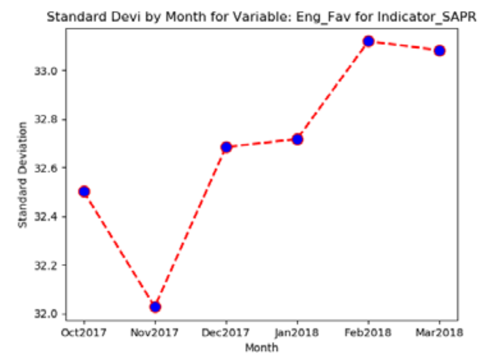

Mean by Month for Variable: JobSat_Fav for Indicator_SAPR
Mean
69.00
68.75
68.50
68.25
68.00
67.75
67.50
67.25
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

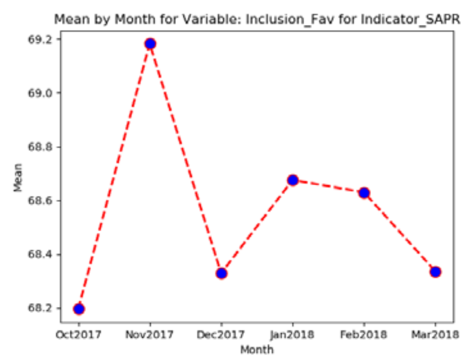

Standard Devi by Month for Variable: JobSat_Fav for Indicator_SAPR
Standard Deviation
44.4
44.3
44.2
44.1
44.0
43.9
43.8
43.7
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

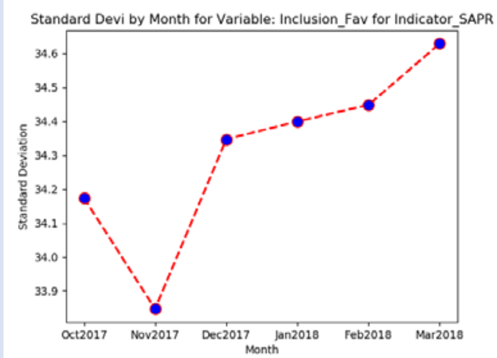

Mean by Month for Variable: OrgProc_Fav for Indicator_SAPR
Mean
66.2
66.0
65.8
65.6
65.4
65.2
65.0
64.8
64.6
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

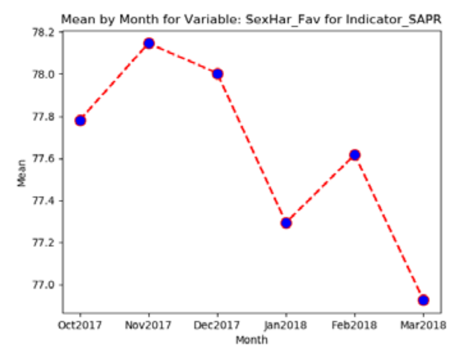

Standard Devi by Month for Variable: OrgProc_Fav for Indicator_SAPR
Standard Deviation
39.7
39.6
39.5
39.4
39.3
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

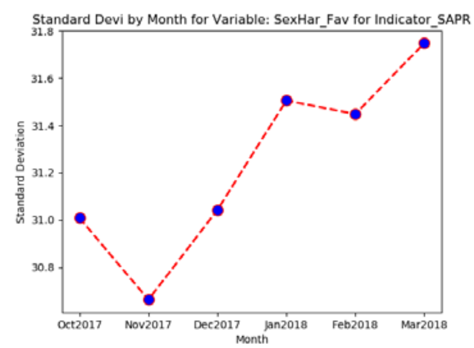

Mean by Month for Variable: Eng_Fav for Indicator_SAPR
Mean
78.0
77.8
77.6
77.4
77.2
77.0
76.8
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

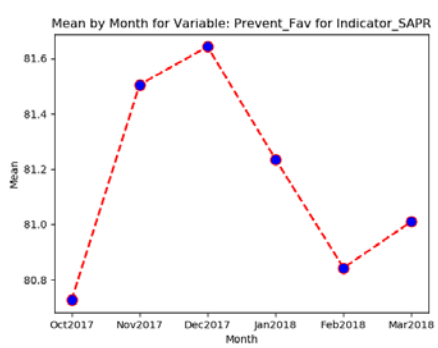

Standard Devi by Month for Variable: Eng_Fav for Indicator_SAPR
Standard Deviation
33.0
32.8
32.6
32.4
32.2
32.0
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

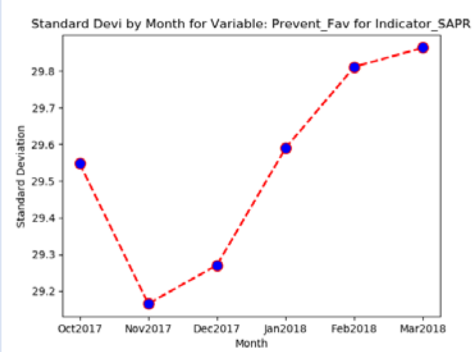

Mean by Month for Variable: Inclusion_Fav for Indicator_SAPR
Mean
69.2
69.0
68.8
68.6
68.4
68.2
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

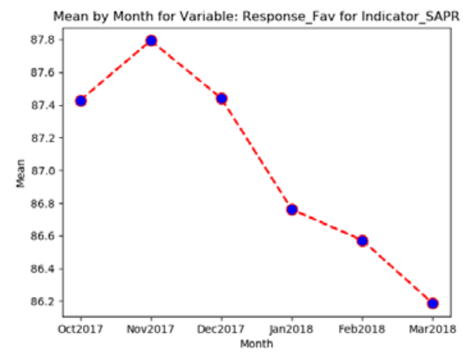

Standard Devi by Month for Variable: Inclusion_Fav for Indicator_SAPR
Standard Deviation
34.6
34.5
34.4
34.3
34.2
34.1
34.0
33.9
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

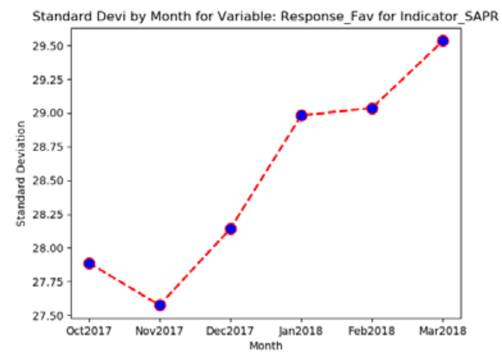

Mean by Month for Variable: SexHar_Fav for Indicator_SAPR
Mean
78.2
78.0
77.8
77.6
77.4
77.2
77.0
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

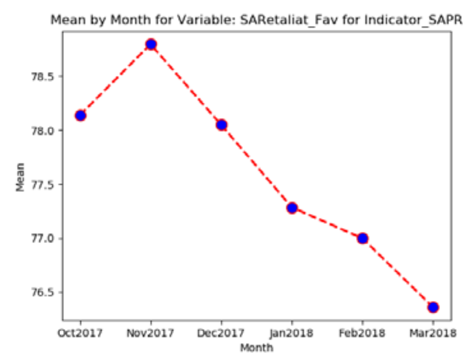

Standard Devi by Month for Variable: SexHar_Fav for Indicator_SAPR
Standard Deviation
31.8
31.6
31.4
31.2
31.0
30.8
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

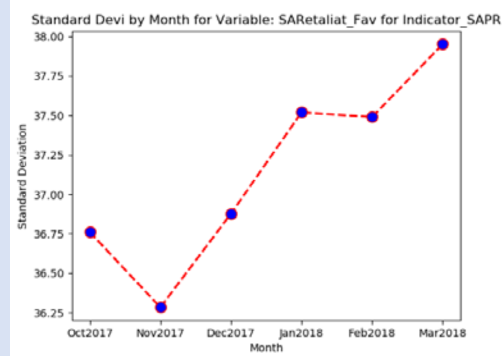

Mean by Month for Variable: Prevent_Fav for Indicator_SAPR
Mean
81.6
81.4
81.2
81.0
80.8
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

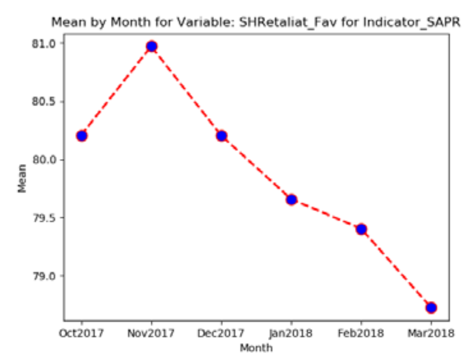

Standard Devi by Month for Variable: Prevent_Fav for Indicator_SAPR
Standard Deviation
29.8
29.7
29.6
29.5
29.4
29.3
29.2
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month


Mean by Month for Variable: Response_Fav for Indicator_SAPR
Mean
87.8
87.6
87.4
87.2
87.0
86.8
86.6
86.4
86.2
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month


Standard Devi by Month for Variable: Response_Fav for Indicator_SAPR
Standard Deviation
29.50
29.25
29.00
28.75
28.50
28.25
28.00
27.75
27.50
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month


Mean by Month for Variable: SARetaliat_Fav for Indicator_SAPR
Mean
78.5
78.0
77.5
77.0
76.5
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month


Standard Devi by Month for Variable: SARetaliat_Fav for Indicator_SAPR
Standard Deviation
38.00
37.75
37.50
37.25
37.00
36.75
36.50
36.25
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month


Mean by Month for Variable: SHRetaliat_Fav for Indicator_SAPR
Mean
81.0
80.5
80.0
79.5
79.0
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

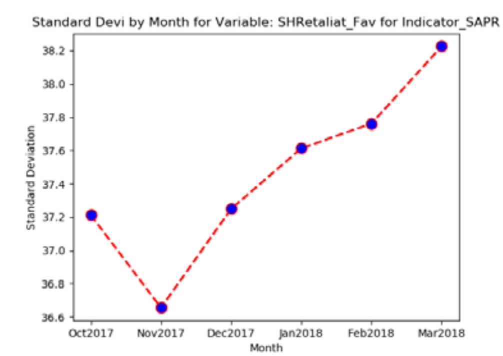
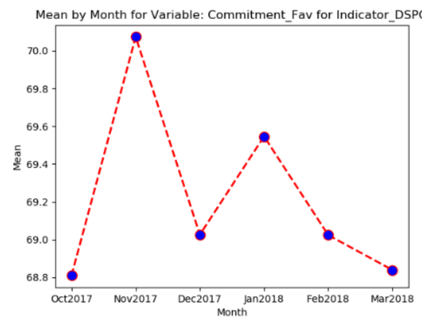
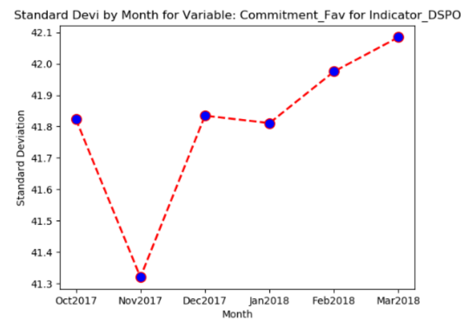
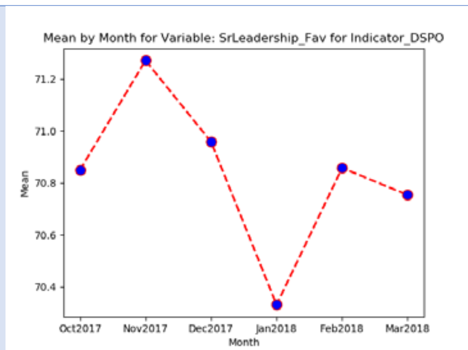

Standard Devi by Month for Variable: SHRetaliat_Fav for Indicator_SAPR
Standard Deviation
38.2
38.0
37.8
37.6
37.4
37.2
37.0
36.8
36.6
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

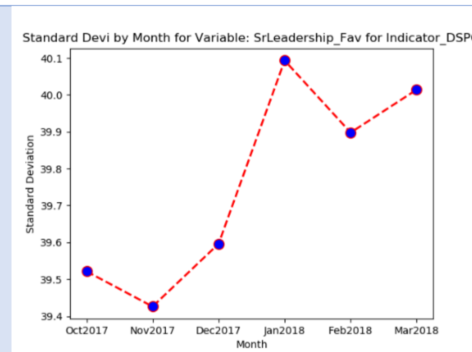
Figure 3: Mean and standard deviation of key survey variables over time period for DSPO

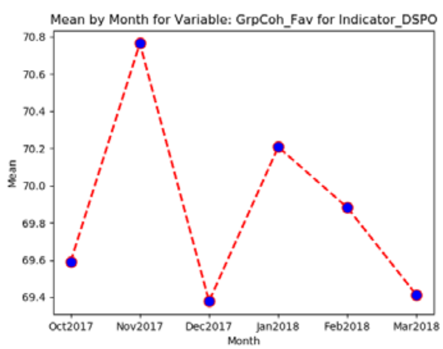
Mean Value for Survey Variable	Standard Deviation for Survey Variable

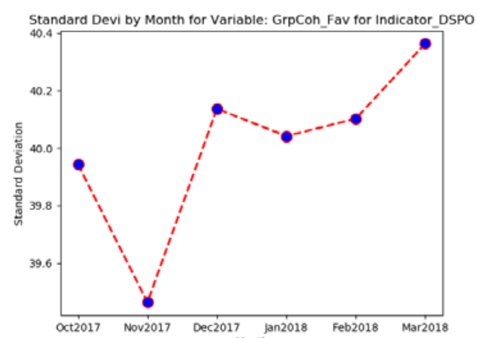


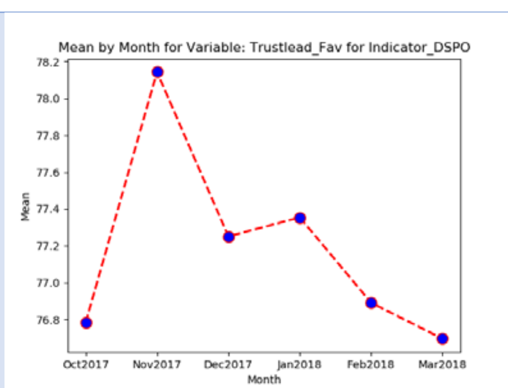


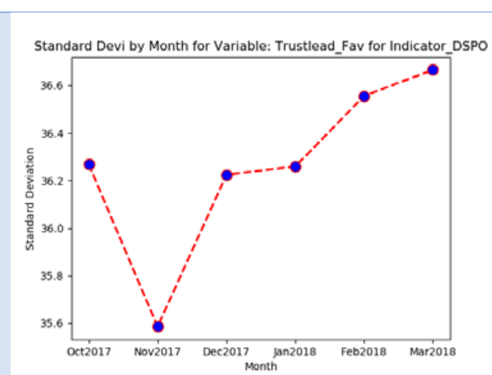


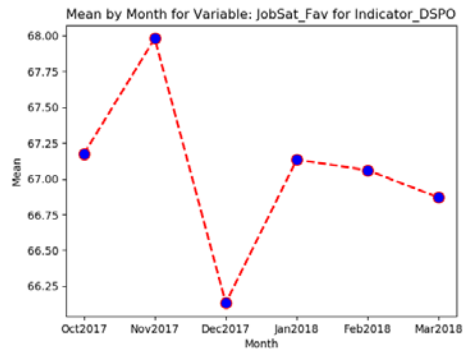


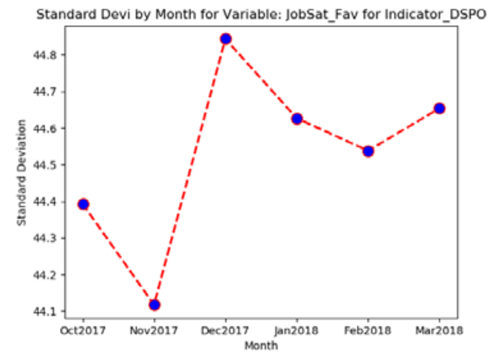


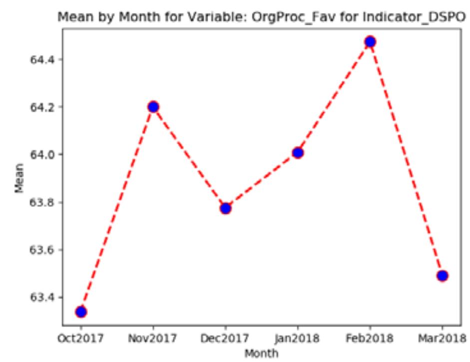


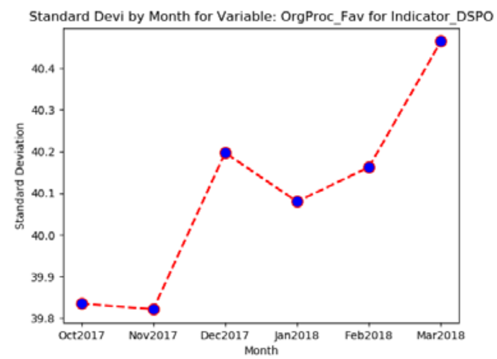


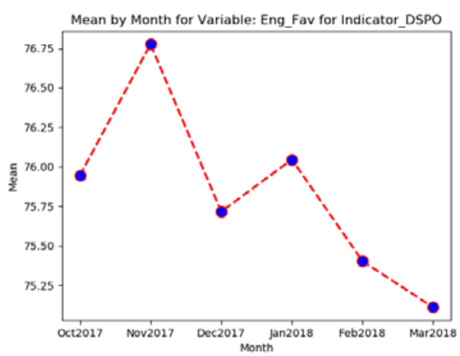


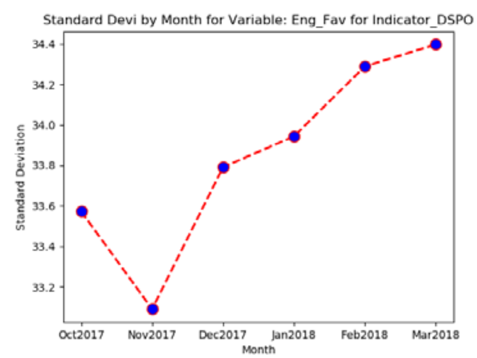

Mean by Month for Variable: JobSat_Fav for Indicator_DSPO
Mean
68.00
67.75
67.50
67.25
67.00
66.75
66.50
66.25
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

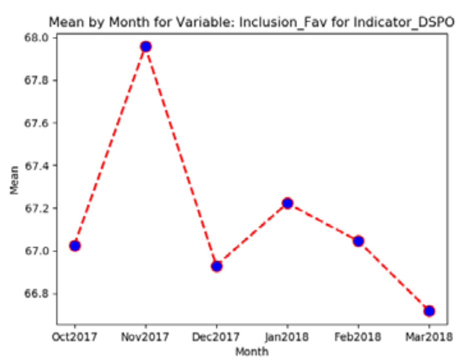

Standard Devi by Month for Variable: JobSat_Fav for Indicator_DSPO
Standard Deviation
44.8
44.7
44.6
44.5
44.4
44.3
44.2
44.1
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

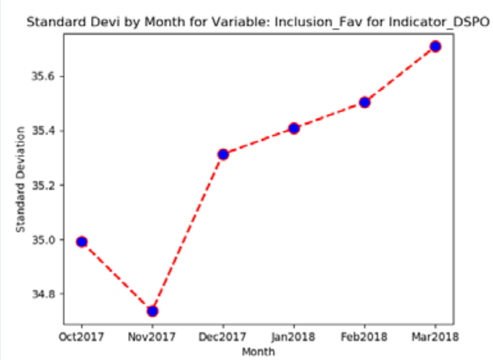

Mean by Month for Variable: OrgProc_Fav for Indicator_DSPO
Mean
64.4
64.2
64.0
63.8
63.6
63.4
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

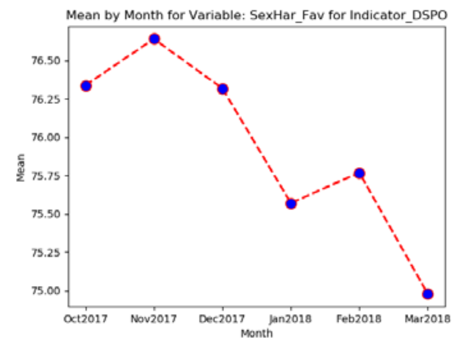

Standard Devi by Month for Variable: OrgProc_Fav for Indicator_DSPO
Standard Deviation
40.4
40.3
40.2
40.1
40.0
39.9
39.8
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

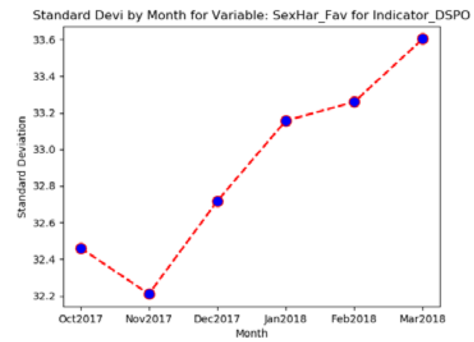

Mean by Month for Variable: Eng_Fav for Indicator_DSPO
Mean
76.75
76.50
76.25
76.00
75.75
75.50
75.25
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

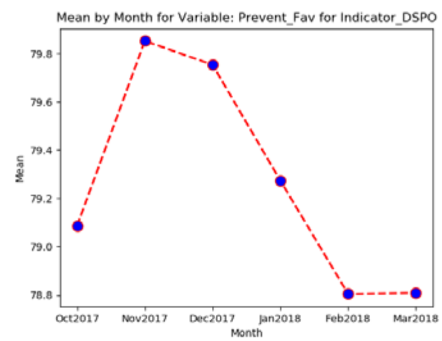

Standard Devi by Month for Variable: Eng_Fav for Indicator_DSPO
Standard Deviation
34.4
34.2
34.0
33.8
33.6
33.4
33.2
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

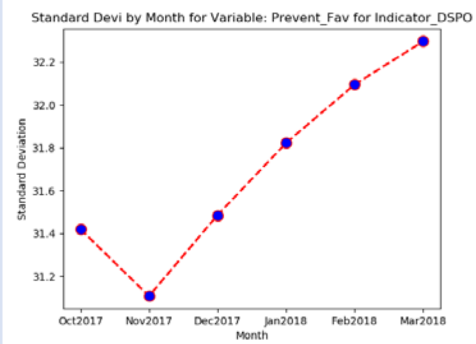

Mean by Month for Variable: Inclusion_Fav for Indicator_DSPO
Mean
68.0
67.8
67.6
67.4
67.2
67.0
66.8
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

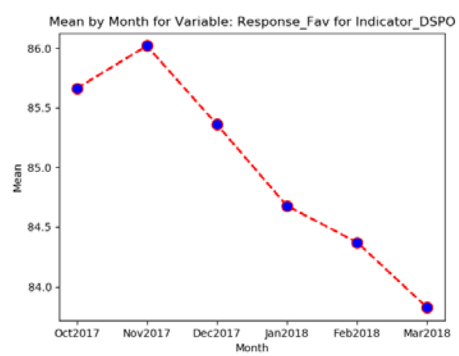

Standard Devi by Month for Variable: Inclusion_Fav for Indicator_DSPO
Standard Deviation
35.6
35.4
35.2
35.0
34.8
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

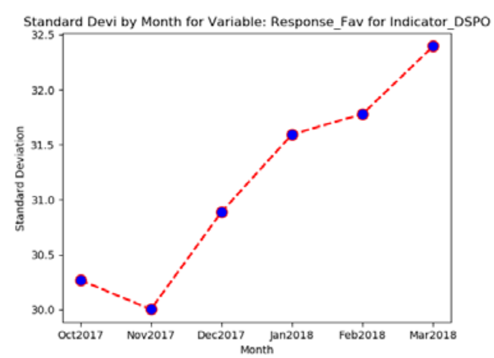

Mean by Month for Variable: SexHar_Fav for Indicator_DSPO
Mean
Month
76.50
76.25
76.00
75.75
75.50
75.25
75.00
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018

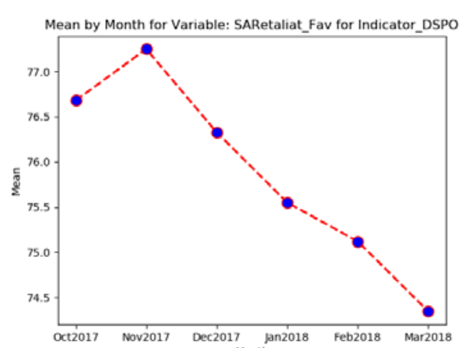

Standard Devi by Month for Variable: SexHar_Fav for Indicator_DSPO
Standard Deviation
Month
33.6
33.4
33.2
33.0
32.8
32.6
32.4
32.2
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018

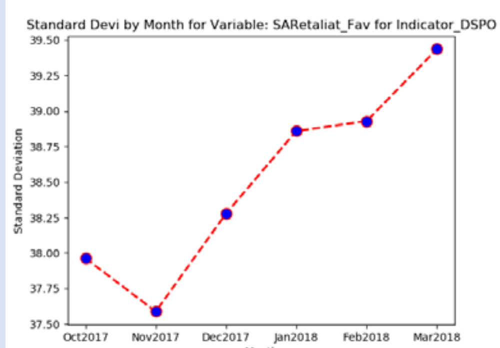

Mean by Month for Variable: Prevent_Fav for Indicator_DSPO
Mean
Month
79.8
79.6
79.4
79.2
79.0
78.8
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018

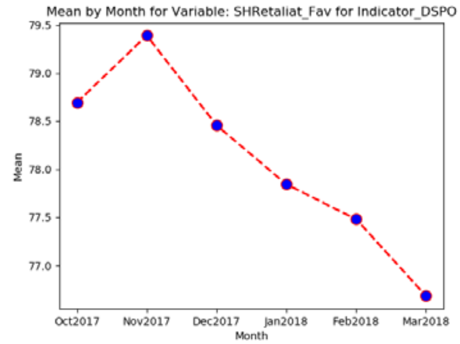

Standard Devi by Month for Variable: Prevent_Fav for Indicator_DSPO
Standard Deviation
Month
32.2
32.0
31.8
31.6
31.4
31.2
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018


Mean by Month for Variable: Response_Fav for Indicator_DSPO
Mean
Month
86.0
85.5
85.0
84.5
84.0
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018


Standard Devi by Month for Variable: Response_Fav for Indicator_DSPO
Standard Deviation
Month
32.5
32.0
31.5
31.0
30.5
30.0
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018


Mean by Month for Variable: SARetaliat_Fav for Indicator_DSPO
Mean
Month
77.0
76.5
76.0
75.5
75.0
74.5
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018


Standard Devi by Month for Variable: SARetaliat_Fav for Indicator_DSPO
Standard Deviation
Month
39.50
39.25
39.00
38.75
38.50
38.25
38.00
37.75
37.50
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018



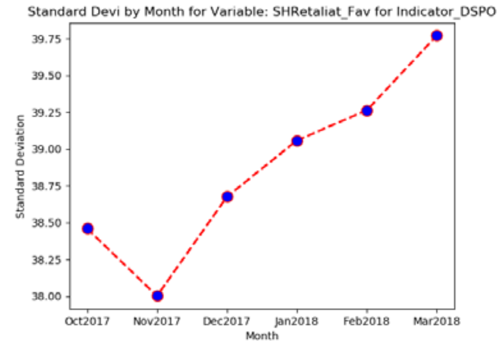
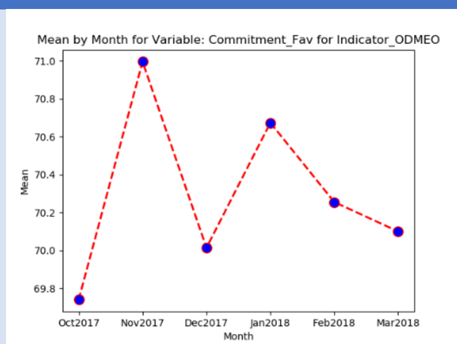
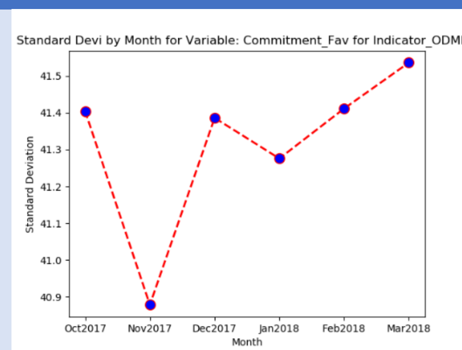
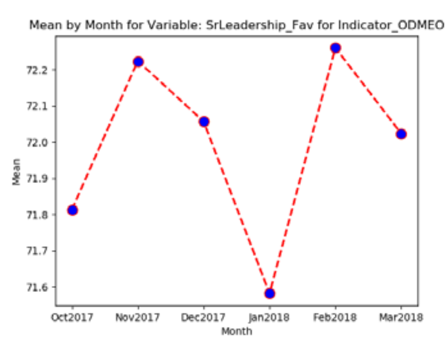
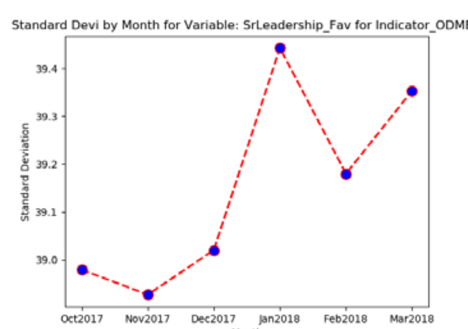
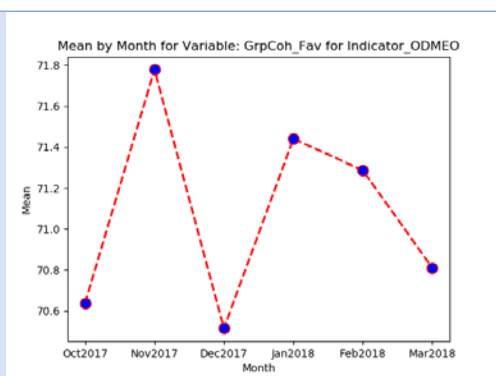
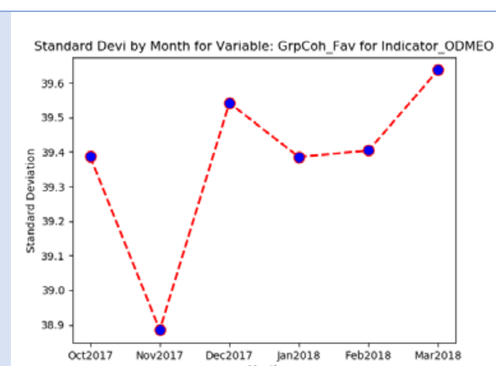
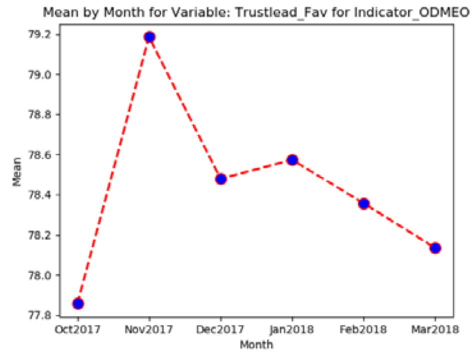
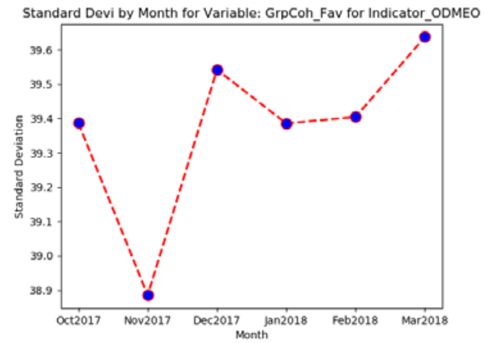
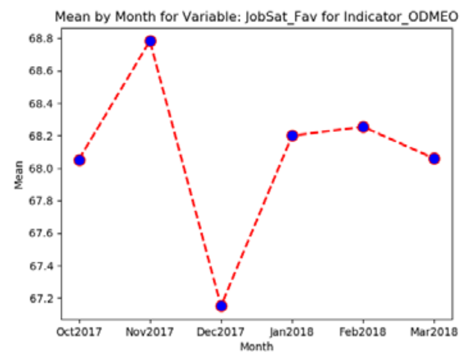


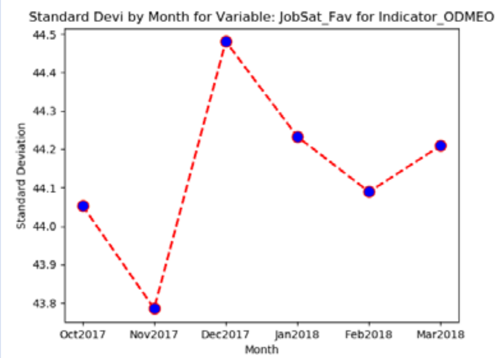
Figure 4: Mean and standard deviation of key survey variables over time period for ODMEO

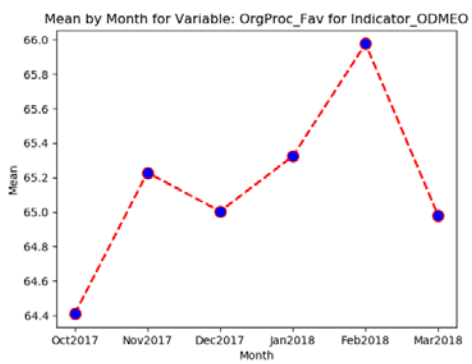
Mean Value for Survey Variable	Standard Deviation for Survey Variable
	
	
	

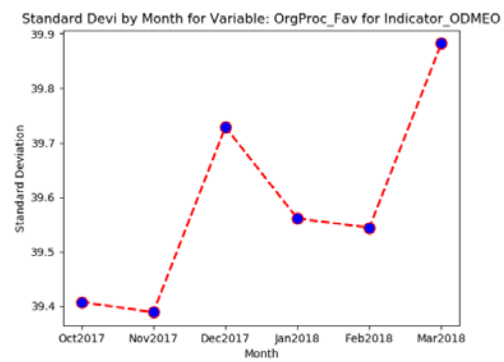

Mean by Month for Variable: Trustlead_Fav for Indicator_ODMEO
Mean
79.2
79.0
78.8
78.6
78.4
78.2
78.0
77.8
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

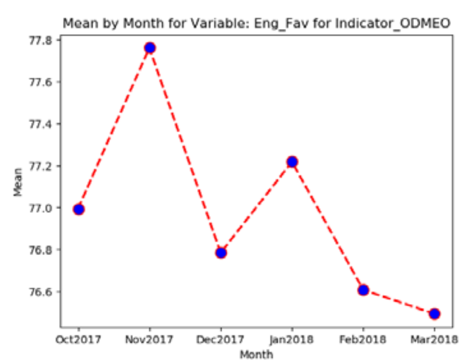

Standard Devi by Month for Variable: GrpCoh_Fav for Indicator_ODMEO
Standard Deviation
39.6
39.5
39.4
39.3
39.2
39.1
39.0
38.9
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

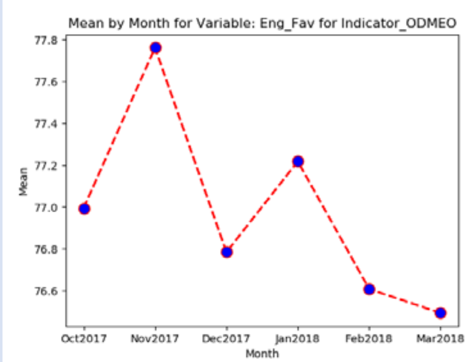

Mean by Month for Variable: JobSat_Fav for Indicator_ODMEO
Mean
68.8
68.6
68.4
68.2
68.0
67.8
67.6
67.4
67.2
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

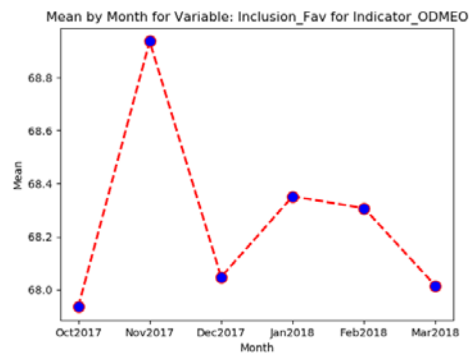

Standard Devi by Month for Variable: JobSat_Fav for Indicator_ODMEO
Standard Deviation
44.5
44.4
44.3
44.2
44.1
44.0
43.9
43.8
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

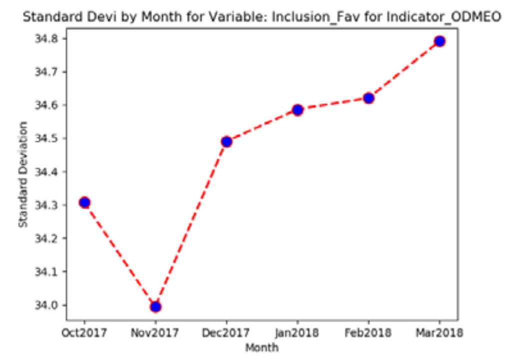

Mean by Month for Variable: OrgProc_Fav for Indicator_ODMEO
Mean
66.0
65.8
65.6
65.4
65.2
65.0
64.8
64.6
64.4
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

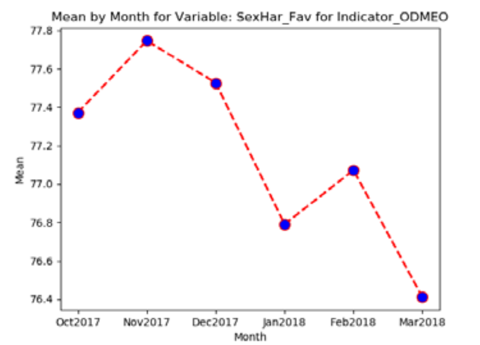

Standard Devi by Month for Variable: OrgProc_Fav for Indicator_ODMEO
Standard Deviation
39.9
39.8
39.7
39.6
39.5
39.4
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

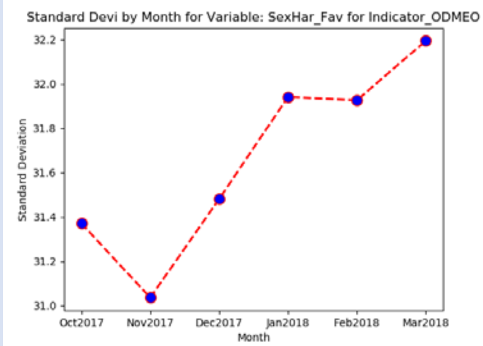

Mean by Month for Variable: Eng_Fav for Indicator_ODMEO
Mean
77.8
77.6
77.4
77.2
77.0
76.8
76.6
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

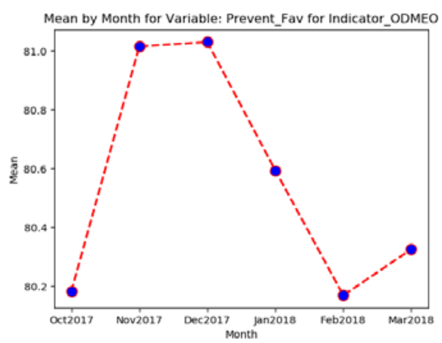

Mean by Month for Variable: Eng_Fav for Indicator_ODMEO
Mean
77.8
77.6
77.4
77.2
77.0
76.8
76.6
Oct2017 Nov2017 Dec2017 Jan2018 Feb2018 Mar2018
Month

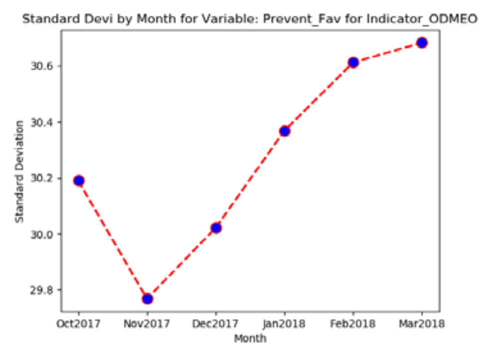

Mean by Month for Variable: Inclusion_Fav for Indicator_ODMEO
Mean
68.0
68.2
68.4
68.6
68.8
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

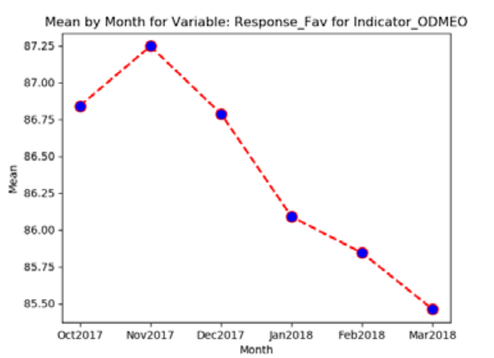

Standard Devi by Month for Variable: Inclusion_Fav for Indicator_ODMEO
Standard Deviation
34.0
34.1
34.2
34.3
34.4
34.5
34.6
34.7
34.8
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

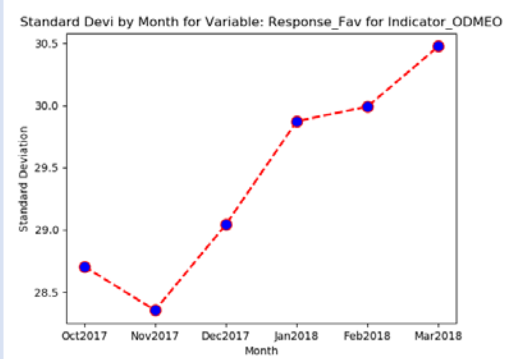

Mean by Month for Variable: SexHar_Fav for Indicator_ODMEO
Mean
76.4
76.6
76.8
77.0
77.2
77.4
77.6
77.8
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month

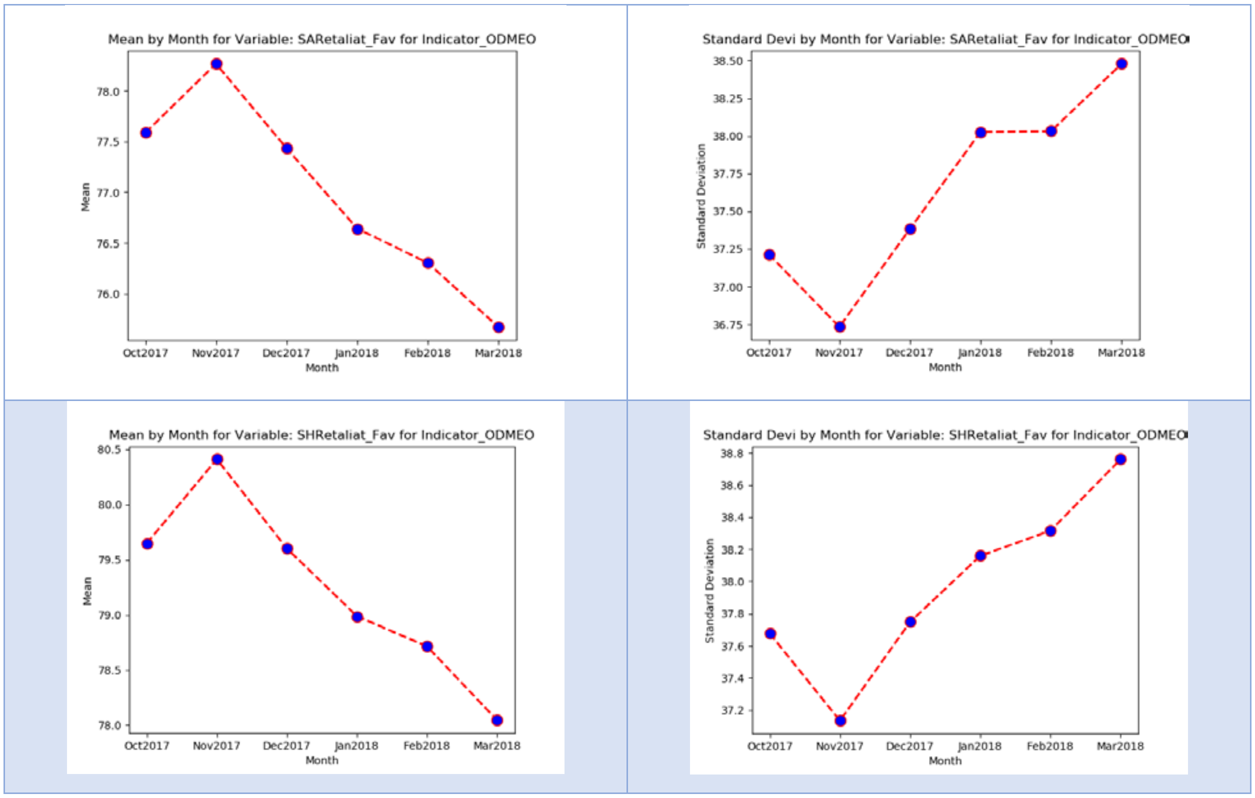

Standard Devi by Month for Variable: SexHar_Fav for Indicator_ODMEO
Standard Deviation
31.0
31.2
31.4
31.6
31.8
32.0
32.2
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month


Mean by Month for Variable: Prevent_Fav for Indicator_ODMEO
Mean
80.2
80.4
80.6
80.8
81.0
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month


Standard Devi by Month for Variable: Prevent_Fav for Indicator_ODMEO
Standard Deviation
29.8
30.0
30.2
30.4
30.6
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month


Mean by Month for Variable: Response_Fav for Indicator_ODMEO
Mean
85.50
85.75
86.00
86.25
86.50
86.75
87.00
87.25
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month


Standard Devi by Month for Variable: Response_Fav for Indicator_ODMEO
Standard Deviation
28.5
29.0
29.5
30.0
30.5
Oct2017
Nov2017
Dec2017
Jan2018
Feb2018
Mar2018
Month



7. Unequal Weighting Effect

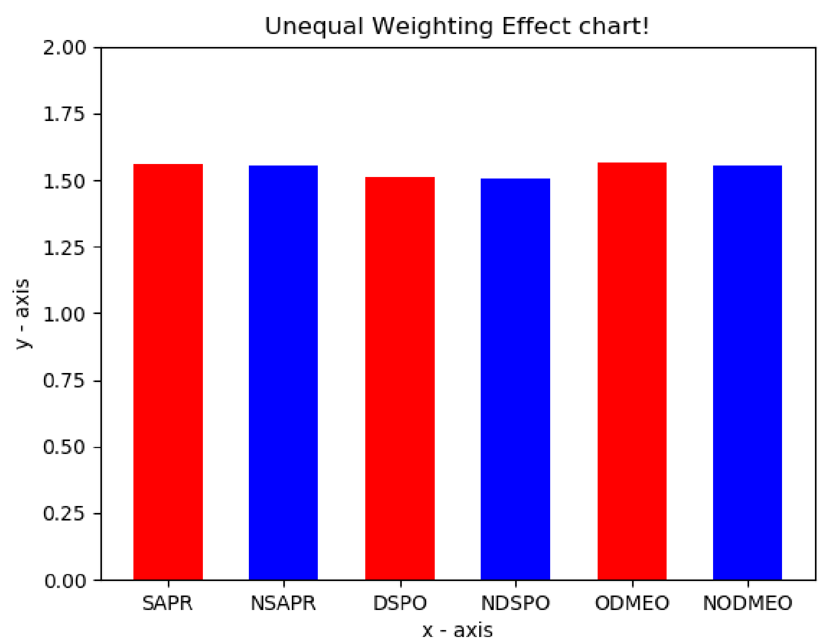
When variation in the weights between different time period or units are large, weighting trimming is necessary. Thus, study of the unequal weighting effect (UWE) is necessary. UWE is an upper bound of the variance ratio of an estimate, calculated from a survey to the variance one would obtain from a simple random sample with the same sample size (Biemer and Christ, 2008; Office on Smoking and Health & Office of Science, 2014) give the concept of UWE: Kish (1965) derived a formula for determining the maximum increase in variance of an estimate of a population mean due to a weight variation. His formula assumes there is no correlation between the survey weights and the characteristic whose mean is to be estimated. This may be a good approximation for many survey variables because the survey design and weight adjustments are optimized for only a few key characteristics out of hundreds that may be collected in a survey. The actual variance increase will vary across characteristics in the survey and will be smaller for characteristics where the covariance between the observations and the weights are larger. Under these assumptions, Kish obtained the following expression for the UWE defined as the ratio of the variances of the weighted mean to the variance of the unweighted mean:

$$\begin{aligned} UWE &= 1 + cv(w)^2 \\ &= \frac{n\sum_{i=1}^{n} w_i^2}{(\sum_{i=1}^{n} w_i)^2} \end{aligned}$$

where $cv(w)$ is the coefficient of variance of the weights or the sample standard deviation of the weights divided by the sample average weight. From Table 19, the UWE for all three weights are less than 2, so weight trimming will not be applied.

Table 19. Unequal Weighting Effect

Weights	NonRespWeightCHAID_SAPR	NormedWeightSAPR	NonRespWeightCHAID_DSPO	NormedWeightDSPO	NonRespWeightCHAID_ODMEO	NormedWeightODMEO
UWE	1.56039	1.55330	1.51081	1.50720	1.56226	1.55470



8. Python Syntax

8.1 Python Syntax for Algorithm 1 Realized in SPSS:

Syntax1_06082018_SAPR_ver2.sps   Syntax1_06082018_DSPO_ver2.sps   Syntax1_06082018_ODMEO_ver2.sps

```
* Encoding: UTF-8.
OUTPUT CLOSE ALL.
GET FILE="E:/WEI WAN/My SPSS/Summer2018/WorkingData/dataset1_indicatorSAPRequals1.sav".
BEGIN PROGRAM.
import spss,spssaux

vars = ["Commitment_Fav", "SrLeadership_Fav","OrgPerf_Fav", "GrpCoh_Fav","Trustlead_Fav", \
 "JobSat_Fav", "OrgProc_Fav","Eng_Fav", "Inclusion_Fav", \
"SexHar_Fav", "Prevent_Fav", "Response_Fav", "SARetaliat_Fav","SHRetaliat_Fav"]

vector_bias=dict()
temp_list = list()
for i in vars:
    print("This is variable: {}".format(i))
    spss.Submit(r"""
    FREQUENCIES VARIABLES=%(i)s
     /ORDER=ANALYSIS.
    DESCRIPTIVES VARIABLES=%(i)s
     /STATISTICS=MEAN SUM STDDEV VARIANCE RANGE MIN MAX KURTOSIS SKEWNESS.

    GGRAPH
     /GRAPHDATASET NAME="graphdataset" VARIABLES=%(i)s MISSING=LISTWISE REPORTMISSING=N
     /GRAPHSPEC SOURCE=INLINE.
    BEGIN GPL
     SOURCE: s=userSource(id("graphdataset"))
     DATA: %(i)s=col(source(s), name("%(i)s"))
     GUIDE: axis(dim(1), label("%(i)s"))
     GUIDE: axis(dim(2), label("Frequency Percent"))
     ELEMENT: interval(position(summary.percent.count(bin.rect(%(i)s),
       base.all(acrossPanels()))), shape.interior(shape.square))
    END GPL.      """ %locals())

    cmd="DESCRIPTIVES VARIABLES={}.".format(i)
    desc_table,errcode=spssaux.CreateXMLOutput(
                  cmd,
                  omsid="Descriptives")
    meansal=spssaux.GetValuesFromXMLWorkspace(
            desc_table,
            tableSubtype="Descriptive Statistics",
            rowCategory="{}".format(i),
            colCategory="Mean",
            cellAttrib="number")
    if meansal:
        temp1= float(meansal[0].encode('ascii','ignore'))
        print("The mean {} is:".format(i), temp1)

    spss.Submit(r"""
    WEIGHT BY NonRespWeightCHAID_SAPR.

    FREQUENCIES VARIABLES=%(i)s
     /ORDER=ANALYSIS.
    DESCRIPTIVES VARIABLES=%(i)s
     /STATISTICS=MEAN SUM STDDEV VARIANCE RANGE MIN MAX KURTOSIS SKEWNESS.

    GGRAPH
     /GRAPHDATASET NAME="graphdataset" VARIABLES=%(i)s MISSING=LISTWISE REPORTMISSING=N
     /GRAPHSPEC SOURCE=INLINE.
    BEGIN GPL
```

8.2 Python Syntax for Algorithm 1 Realized in SPSS:

Syntax_06212018_ver10NormWeights_Service.sps

```
* Encoding: UTF-8. Syntax_06212018_ver10NormWeights_Service. Important one !.
* Compute effects of weights at the same time.
* Compute ratio for normalization of weights.
* compute normalized weights.
* this file works for all reports.
output close all.
BEGIN PROGRAM.
import spss

spss.Submit(r"""
get file='E:/WEI WAN/My SPSS/Summer2018/WorkingData/01. Cumulative with NR weights October 2017 -
March 2018_For Dr. Wan.sav'.
DATASET NAME alldata.
""")

groupVar = "a_DEOCSID7"
RealTotalVar = "TotalAdminRequested_Replaced"
sumVar1 = "NonRespWeightCHAID_SAPR"
sumVar2 = "NonRespWeightCHAID_DSPO"
sumVar3 = "NonRespWeightCHAID_ODMEO"
flagVar = "a_Service"

#Determine variable indexes from variable names
varCount = spss.GetVariableCount()
groupIndex = 0
RealTotalIndex = 0
sumIndex1 = 0
sumIndex2 = 0
sumIndex3 = 0
flagIndex = 0

for i in range(varCount):
    varName = spss.GetVariableName(i)
    if varName == groupVar:
        groupIndex = i
        continue
    elif varName == RealTotalVar:
        RealTotalIndex = i
        continue
    elif varName == sumVar1:
        sumIndex1 = i
        continue
    elif varName == sumVar2:
        sumIndex2 = i
        continue
    elif varName == sumVar3:
        sumIndex3 = i
        continue
    elif varName == flagVar:
        flagIndex = i
        continue

varIndex = [groupIndex, RealTotalIndex, sumIndex1,  sumIndex2, sumIndex3, flagIndex]
#-----------------------------------------------------------------------#
cur = spss.Cursor(varIndex)
Counts={}
Statistic1={};Statistic2={};Statistic3={}
diff1={}; diff2={}; diff3={}
```

8.3 Python Syntax for Algorithm 3 Realized in SPSS:

Syntax1_06222018_ver1.sps

```
* Encoding: UTF-8. Important one.
* Generate bias ratio for all reports. Outputs are three indepdent files.
OUTPUT CLOSE ALL.
BEGIN PROGRAM.
import spss,spssaux

spss.Submit(r"""
GET FILE="E:\WEI WAN\My
SPSS\Summer2018\WorkingData/dataset1_AllDatasetwithNormedWeights1.sav".
DATASET NAME alldata.
""")

ListofIndicator = ["Indicator_SAPR", "Indicator_ODMEO", "Indicator_DSPO"]
vars = ["Commitment_Fav", "SrLeadership_Fav","OrgPerf_Fav", "GrpCoh_Fav","Trustlead_Fav", \
 "JobSat_Fav", "OrgProc_Fav","Eng_Fav", "Inclusion_Fav", \
"SexHar_Fav", "Prevent_Fav", "Response_Fav", "SARetaliat_Fav","SHRetaliat_Fav"]


for k in ListofIndicator:
    print("This is for {} report".format(k))
    spss.Submit(r"""
                DATASET ACTIVATE alldata.
                DATASET COPY  normed_%(k)s.
                DATASET ACTIVATE  normed_%(k)s.
                FILTER OFF.
                USE ALL.
                SELECT IF (%(k)s = 1).
                EXECUTE.
                """ %locals())
#-------------------------------------------------------------------------------------------
    vector_bias=dict()
    temp_list = list()
    for i in vars:
        print("This is variable: {}".format(i))
        spss.Submit(r"""
        FREQUENCIES VARIABLES=%(i)s
          /ORDER=ANALYSIS.
        DESCRIPTIVES VARIABLES=%(i)s
          /STATISTICS=MEAN SUM STDDEV VARIANCE RANGE MIN MAX KURTOSIS SKEWNESS.

        GGRAPH
          /GRAPHDATASET NAME="graphdataset" VARIABLES=%(i)s MISSING=LISTWISE REPORTMISSING
          /GRAPHSPEC SOURCE=INLINE.
        BEGIN GPL
          SOURCE: s=userSource(id("graphdataset"))
          DATA: %(i)s=col(source(s), name("%(i)s"))
          GUIDE: axis(dim(1), label("%(i)s"))
          GUIDE: axis(dim(2), label("Frequency Percent"))
          ELEMENT: interval(position(summary.percent.count(bin.rect(%(i)s),
            base.all(acrossPanels()))), shape.interior(shape.square))
        END GPL.      """ %locals())

        cmd="DESCRIPTIVES VARIABLES={}.".format(i)
        desc_table,errcode=spssaux.CreateXMLOutput(
                        cmd,
                        omsid="Descriptives")
        meansal=spssaux.GetValuesFromXMLWorkspace(
                desc_table,
```

8.4 Python Syntax for Algorithm 4 Realized in SPSS:

Syntax1_07092018_ver6.sps

```
* Encoding: UTF-8. Important one.
* Successfully iterate different reports and iterate in all variables.
* Successfully generate data file for each variables.
* Successfuly get z-score.
OUTPUT CLOSE ALL.
BEGIN PROGRAM.
import spss,spssaux
import matplotlib.pyplot as plt
import numpy as np
from pylab import figure, axes, pie, title, show

spss.Submit(r"""
GET FILE="E:\WEI WAN\My
SPSS\Summer2018\WorkingData/dataset1_AllDatasetwithNormedWeights1.sav".
DATASET NAME alldata.
""")

ListofIndicator = ["Indicator_SAPR", "Indicator_ODMEO", "Indicator_DSPO"]
vars = ["Commitment_Fav", "SrLeadership_Fav","OrgPerf_Fav", "GrpCoh_Fav","Trustlead_Fav", \
 "JobSat_Fav", "OrgProc_Fav","Eng_Fav", "Inclusion_Fav", \
"SexHar_Fav", "Prevent_Fav", "Response_Fav", "SARetaliat_Fav","SHRetaliat_Fav"]
list_month = ["Oct2017", "Nov2017", "Dec2017", "Jan2018", "Feb2018", "Mar2018"]
length = len(list_month)

for m in ListofIndicator:
    print("This is for {} report".format(m))
    spss.Submit(r"""
                DATASET ACTIVATE alldata.
                DATASET COPY  normed_%(m)s.
                DATASET ACTIVATE  normed_%(m)s.
                FILTER OFF.
                USE ALL.
                SELECT IF (%(m)s = 1).
                EXECUTE.
                """ %locals())

    tempweightvar = "NormedWeight"+m.lstrip('Indicator_')
    spss.Submit(r"""
                 WEIGHT BY %(tempweightvar)s.
                 """%locals())

    mean_dic = dict()
    var_dic = dict()
    for k in vars:
        cmd="MEANS TABLES={} BY Month  /CELLS=MEAN STDDEV.".format(k)
        desc_table,errcode=spssaux.CreateXMLOutput(
                cmd,
                omsid="Means")

        monthmean=spssaux.GetValuesFromXMLWorkspace(
              desc_table,
              tableSubtype="Report",
              colCategory="Mean",
              cellAttrib="number")

        monthvar=spssaux.GetValuesFromXMLWorkspace(
              desc_table,
              tableSubtype="Report",
```

8.5 Python Syntax for Algorithm 5 Realized in SPSS:

Syntax1_07112018_ver1.sps

```
* Encoding: UTF-8.
* Computer UWE.

output close all.
BEGIN PROGRAM.
import spss
import matplotlib.pyplot as plt

spss.Submit(r"""
get file='E:/WEI WAN/My SPSS/Summer2018/WorkingData/dataset1_AllDatasetwithNormedWeights1.sav'.
DATASET NAME alldata.
""")

weight_SAPR = "NonRespWeightCHAID_SAPR"
weight_DSPO = "NonRespWeightCHAID_DSPO"
weight_ODMEO = "NonRespWeightCHAID_ODMEO"

Nweight_SAPR = "NormedWeightSAPR"
Nweight_DSPO = "NormedWeightDSPO"
Nweight_ODMEO = "NormedWeightODMEO"

#Determine variable indexes from variable names
varCount = spss.GetVariableCount()

index_weight_SAPR = "NonRespWeightCHAID_SAPR"
index_weight_DSPO = "NonRespWeightCHAID_DSPO"
index_weight_ODMEO = "NonRespWeightCHAID_ODMEO"

index_Nweight_SAPR = "NormedWeightSAPR"
index_Nweight_DSPO = "NormedWeightDSPO"
index_Nweight_ODMEO = "NormedWeightODMEO"


for i in range(varCount):
    varName = spss.GetVariableName(i)
    if varName == weight_SAPR:
        index_weight_SAPR = i
        continue
    elif varName == weight_DSPO:
        index_weight_DSPO = i
        continue
    elif varName == weight_ODMEO:
        index_weight_ODMEO = i
        continue
    elif varName == Nweight_SAPR:
        index_Nweight_SAPR = i
        continue
    elif varName == Nweight_DSPO:
        index_Nweight_DSPO = i
        continue
    elif varName == Nweight_ODMEO:
        index_Nweight_ODMEO = i
        continue

varIndex = [index_weight_SAPR, index_weight_DSPO, index_weight_ODMEO, index_Nweight_SAPR ,
index_Nweight_DSPO, index_Nweight_ODMEO]
cur = spss.Cursor(varIndex)
TotalCase = cur.GetCaseCount()
```

9. References

Baker, H., Fuller, W. A., & McLoughlin, M. (1994). Regression Weighting in the Presence of Nonresponse with Application to the 1987-1988 Nationwide Food Consumption Survey. *Survey Methodology vol. 20,* pp. 75-85.

Bethlehem. J. (1988). Reduction of Nonresponse Bias through Regression Estimation. *Journal of Official Statistics, vol. 4*, pp. 251-260.

Bethlehem, J., & Keller, W. J. (1987). Linear Weighting of Sample Survey Data. *Journal of Official Statistics, vol. 3*, pp. 141-153.

Biemer, P. & Christ, S. (2008). "Weighting Survey Data," in International Handbook survey methodology, Chicago, pp. 317-341.

Brick, J. M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research, vol. 5,* pp. 215 - 238.

Cochran., W. G. (2015). Sampling Techniques, 3rd ed., Wiley India.

Deville. J.-C., & Sarndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association, vol. 87*, pp. 376-382.

Deville J.-C., Sarndal, C.-E., & Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling,. *Journal of the American Statistical Association, vol. 88*, pp. 1013-1020.

Earp, M., Kott, P., Kreuter, F., Mitchell, M., & Porter E. (2012). Nonresponse Bias Adjustment in Establishment Surveys: A Comparison of Weighting Methods using the Agricultural Resource Management Survey. *Section on Government Statistics - JSM 2012*, pp. 2070-2084.

Flores-Cervantes, I., & Kalton, G. (2003). Weighting Methods. *Journal of Official Statistics, vol. 19,* no. 2, pp. 81-97.

Fuller, W.A. (2002). Regression Estimation for Survey Samples. *Survey Methodology, vol. 28*, pp. 5-23. 9

Fulsom, R. E., Iannacchione, V. G., & Milne. J. G. (1991). Response Probability Weight Adjustments Using Logistic Regression. *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 637-642.

Ireland C.T., & Kullback, S. (1968). Contingency Tables With Given Marginals. *Biometrika, vol. 55,* pp. 179-188.

Kalton, G., & Kasprzyk, D. (1986). Treatment of Missing Survey Data. *Survey Methodology, no. 12,* pp. 1-16.

Kalton G., & Maligalig, D. (1991). A Comparison of Methods of Weighting Adjustment for Nonresponse, 1994 Annual Research Conference: March 17-20, 1991: Proceedings: Bureau of the Census.

Kish, L. (1965). Survey Sampling, Wiley-Interscience Publication, New York.

Kish, L. (1992). Weighting for Unequal Pi. *Journal of Official Statistics, vol. 8,* pp. 183-200.

Lepkowski, J., Kalton, G., Kasprzyk, D., & Bureau, U. (1989). Weighting adjustments for partial nonresponse in the 1984 SIPP panel. *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 296-301.

Little, R. (1986). Survey Nonresponse Adjustments for Estimates of Means. *Journal of the American Statistical Association, vol. 77,* pp. 237-250.

Little, R. & Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means?. *Survey Methodology, vol. 31,* pp. 161-168.

Office on Smoking and Health & Office of Science (2014). *2012-2013 National Adult Tobacco Survey Weighting Specifications*. https://www.cdc.gov/tobacco/data_statistics/surveys/nats/pdfs/2012_13_weighting_methodology.pdf

Phipps, P., & Toth, D. (2012). Analyzing Establishment Nonresponse Using and Interpretable Regression Tree Model with Linked Administrative Data. *Annals of Applied Statistics, vol. 6*, no. 2.

Rosenbaum, P. & Rubin, D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects *Biometrika, vol. 70*, pp. 41-55.