What Score Should Johnny Get? Missing_Items SAS® Macro for Analyzing Missing Item Responses on Summative Scales

Patricia Rodríguez de Gil and Jeffrey D. Kromrey
University of South Florida, Tampa, FL

ABSTRACT

Missing data are usually not the focus of any given study but researchers frequently encounter missing data when conducting empirical research. Missing data for Likert-type response scales, whose items are often combined to make summative scales, are particularly problematic because of the nature of the constructs typically measured, such as attitudes and opinions. This paper provides a SAS macro, written in SAS/IML and SAS/STAT, for imputation of missing item responses that allows estimation of person-level means or sums across items in the scale. Imputations are obtained using multiple imputation (MI), single regression substitution (SRS), relative mean substitution (RMS), and person mean substitution (PMS). In addition, the results of a simulation study comparing the accuracy and precision of the imputation methods are summarized.

Keywords: Missing data, missing items, summative scales

INTRODUCTION

Missing data for Likert-type response scales, whose items are often combined to make summative scales, are particularly problematic due to the nature of constructs typically measured, such as attitudes and opinions (Raaijmakers, 1999). Research using Likert-type scales tends to have missing data for several reasons. When respondents choose not to answer sensitive questions like income level or sexual behaviors, this type of missing data is called item nonresponse (Buhi, Goodson, & Neilands, 2008; Downey & King, 2001; O'Rourke, 2003). In longitudinal studies, missing data may be due to attrition, or sometimes eligible people choose not to participate. This type of missing data is called unit nonresponse (Buhi et al., 2008; Schafer & Olsen, 1998).

The use of summative scales allow researchers to obtain more dependable and valid measures of constructs but the presence of missing data affects results and interpretation of data; The development of methods for the analysis of samples with incomplete data has been an active area of research (Horton & Kleinman, 2007); however, an agreed upon method for dealing with Likert-type missing data has not been clearly established among researchers (Downey & King, 1998; Raaijmakers, 1999). Also, literature addressing different strategies for dealing with missing data in such circumstances is scarce (Shrive, Stuart, Quan, & Ghali, 2006).

Although there will be some research conditions when the application of simple deletion procedures for handling missing data, such as listwise deletion and pairwise deletion, would be appropriate (e.g., large sample size, low percentage of missing data), previous work on missing data in Likert-type scales have reinforced the idea of the inadequacy of handling item nonresponse using these deletion procedures (Beale & Little, 1975; Little, 1978), mostly if the assumption of data missing at random does not hold. The macro presented in this paper imputes missing data using both the substitution of the person mean for missing item responses and the incorporation of the relationships between items into the estimation of missing values.

MISSING DATA TREATMENTS

Four missing data treatments are implemented in the MISSING_ITEMS macro: Multiple imputation (*MI*), single regression substitution (*SRS*), relative mean substitution (RMS), and person mean substitution (*PMS*).

MULTIPLE IMPUTATION (*MI*)

For the multiple imputation procedure, missing values are replaced with the SAS default of 5 plausible values; this means that five replicate data sets with identical observed values across data sets are generated, but the imputed values vary across the data sets. The variability in missing data allows the MI analysis to incorporate the ambiguity associated with missing data and to reflect the uncertainty about relationships when imputing missing values (Buhi et al., 2008; Patrician, 2002).

SINGLE REGRESION SUBSTITUTION (*SRS*)

In single regression substitution, for each missing item, an observed item most highly correlated to the missing variable is used to predict the missing item value. That is, for a participant presenting valid responses to items 1 through $a-1$, but missing the response for item a, the item that correlates most highly with item a is used to predict the missing item response using the sample linear regression equation to make the prediction.

RELATIVE MEAN SUBSTITUTION (*RMS*)

The relative mean substitution method (*RMS*) is designed specifically to estimate missing values for Likert-type scale items. This method estimates missing data using three sources of information: the person mean of the kth respondent for all valid (nonmissing) item scores, the grand mean of all valid item scores of all respondents, and the mean of all valid scores on the ath item, excluding person *k* (Raaijmakers, 1999).

$$X_{ak} = \left(\frac{\frac{\sum_{i=1}^{n} X_{ik}}{\sum_{j=1}^{N} \sum_{i=1}^{n} X_{ij}}}{Nn} \right) \left(\frac{\sum_{j=1}^{N} X_{aj}}{N} \right) ; (j \neq k)$$

Where:

X_{ak} = the estimated value for missing item a for person k,
i = the valid responses to items 1 to n of person k, and
j = the valid N cases of the sample with no missing data excluding person k

PERSON MEAN SUBSTITUTION (*PMS*)

Considering that the items that compose an attitude scale are developed so that they are correlated with each other (Crocker & Algina, 1986), Downey and King (1998), state that it is reasonable to use a composite, such as the mean, of the observed or responded items by a person, to estimate for his or her missing item. This approach, which substitutes the mean of the nonmissing items for person k for person k's missing items, estimates the person mean using the formula below:

$$X_{ak} = \frac{\sum_{i=1}^{n} X_{ik}}{n}$$

Where:

X_{ak} = the estimated value for missing item a for person k,
i = the valid responses to items 1 to n of person k, and
n = nonmissing items for person k

SIMULATION RESEARCH SUMMARY ON MISSING DATA

Rodríguez and Kromrey (2010) conducted a simulation study to investigate the effectiveness of four imputation procedures multiple imputation (*MI*), single mean substitution (*SMS*), relative mean substitution (*RMS*), and person mean substitution (*PMS*), to estimate missing values on items within summative scales by estimating the extent to which these methods produced stable results over replicated data sets. Response data were generated to model 5-point and 7-point Likert-type scales. The combination of the study's six, fully crossed, fixed factors generated 5184 sets, for which 1000 replications were conducted. Table 1 shows the study design factors.

Sample Size		Likert	Shape	Number of Items		r_{12}		Missing Data Mechanism		Missing Person		Missing Items
10			Uniform	5		.10				.20		.20
50		5	Unimodal	10		.30		Random		.40		.40
100		7	Skewed	20		.50		Nonrandom		.60		.60
500				40								
4	x	2	x 3	x 4	x	3	x	2	x	3	x	3

Table 1. Study Factorial Design
Note: Total conditions = 5184

The effectiveness of the missing data treatments was determined by examining the statistical bias and sampling errors in the estimates of person scores, comparing the results of the completely observed data sets with the results

of the imputed item scores. Statistical bias and root mean squared error (RMSE) were the two statistics studied as measures of the effectiveness of the missing data treatments.

The statistical bias was computed as the average difference between the person score with missing data and the person score that would have been obtained had data not been missing and it was obtained from the difference between the results from the data treatment condition and the true statistics computed from the complete data sets.

$$Bias = \frac{\sum(\hat{\theta}_{ij} - \bar{\theta}_{ij})}{n_i n_j}$$

Where,

$\hat{\theta}_{ij}$ = is the estimated score for respondent i in sample j.

$\bar{\theta}_{ij}$ = is the score value for respondent i in sample j if no data were missing.

n_i, n_j = is the number of respondents and number of samples (respectively) in the simulation condition.

The root mean squared error, which combines bias with the estimated sampling error, provided the total error in the estimate of the person scores:

$$RMSE = \sqrt{\frac{\sum(\hat{\theta}_{ihj} - \theta_{ij})^2}{n_i n_j}}$$

BIAS

Figure 1 shows the distribution of statistical bias for the four missing data methods under the condition of data missing at random. As can be observed in the figure, the methods provided similar mean summary results patterns. Means ranged from -0.0002 (PMS) to 0.0043 (SRS). SRS and MI methods show higher positive bias values than PMS and RMS. For MI, The spread of values was greater; its outside values ranged from -.050 to 0.47.
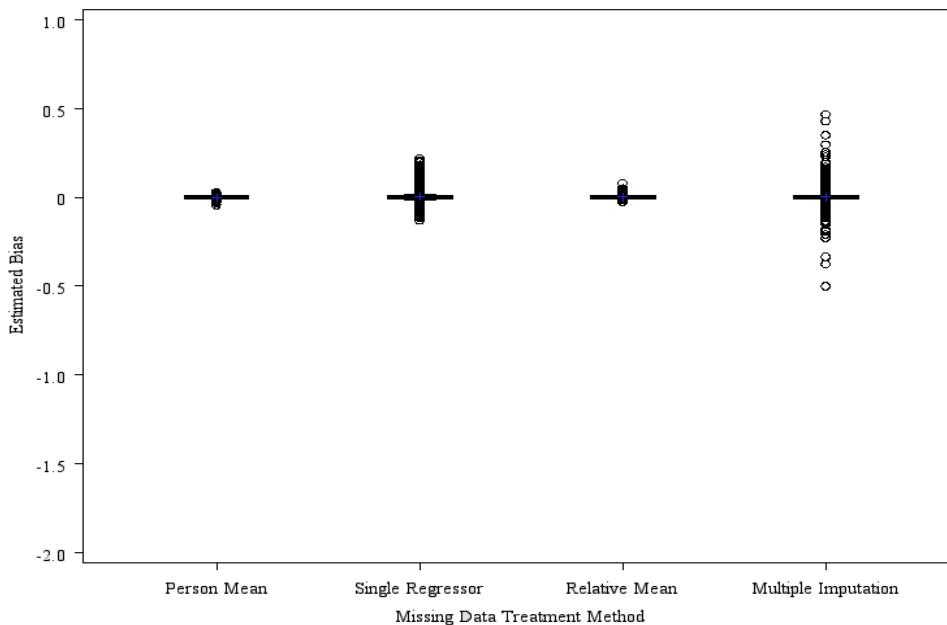
Figure 1. *Distributions of Estimates Statistical Bias in Randomly Missing Data Conditions*

For the nonrandom missing data conditions, Figure 2 shows that the four missing data methods examined provided notably greater bias than that obtained when data were missing at random.

SGF 2013
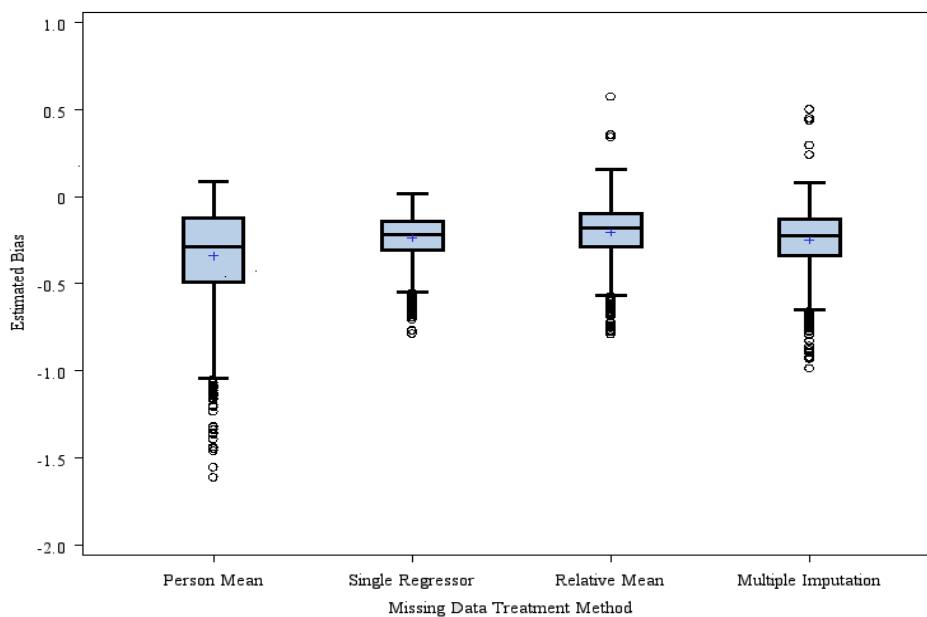
Figure 2. Distributions of Estimates of Statistical Bias in Nonrandom Missing Data Conditions

A comparison of both bias summaries shows that in the nonrandom missing data condition, all four methods led to larger, more extreme biases than those for the random missing data condition (values ranged from .20 for RMS to -0.25 for MI); PMS provided the highest minimum typical value compared to SRS, RMS, and MI. Under the random condition, MI was a particularly poor estimator when sample sizes were very small.

RMSE

The root mean squared error was used also to evaluate the effectiveness of the four missing data treatments in replacing missing data. The RMSE includes both statistical bias and sampling error, providing an index of the total error in a sample estimate. The RMSE summaries for all four missing data treatments under the random and nonrandom condition are shown in Figure 3 and 4 respectively.

Figure 3. Distributions of RMSE for Randomly Missing Data Conditions

In the missing at random condition, all four missing data estimation methods showed outside minimum extreme values that ranged from 0.002 (RMS and SRM) to 0.003 (PMS and MI). A very noticeable observed value was the maximum outside value of 5.0 for the MI procedure. Under nonrandom conditions, all four missing data methods gave high outside positive values that ranged from 1.45 (SRS) to 5.0 (MI), which is indication that all procedures generated large sampling variability; in other words, under the nonrandom condition, the four missing data methods were less effective in estimating the missing values.
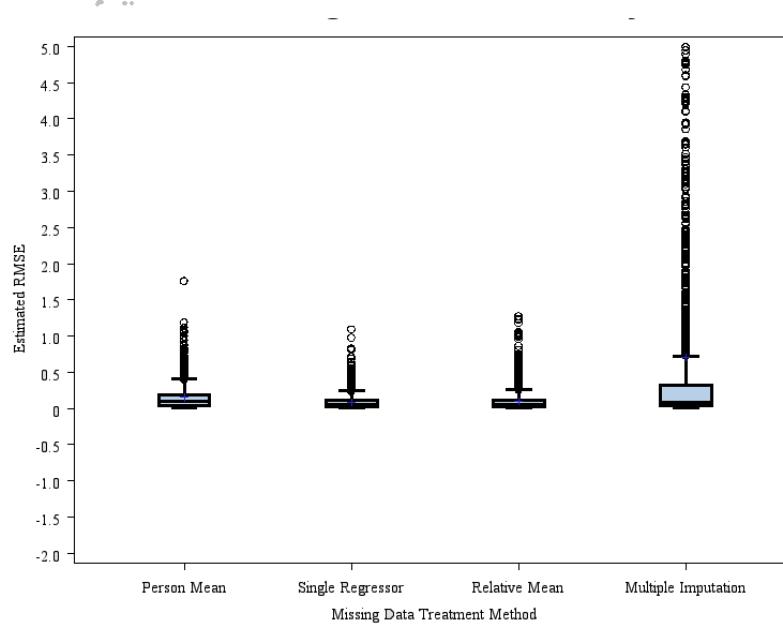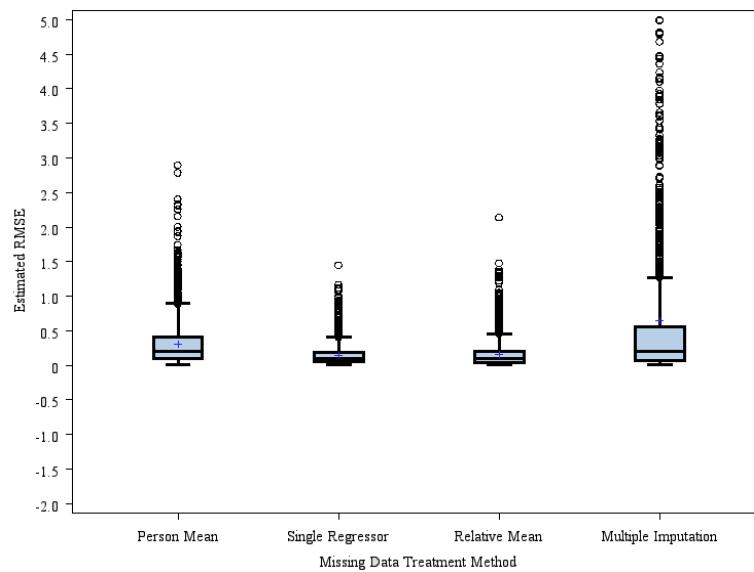

Figure 4. Distributions of RMSE for Nonrandomly Missing Data Conditions

MACRO MISSING_ITEMS

The MISSING_ITEMS macro provides estimates of total scores (either mean or sum) for individual observations. The estimates are obtained using multiple imputation (*MI*), single regression substitution (*SRS*), relative mean substitution (*RMS*), and person mean substitution (*PMS*). The arguments to the macro include the name of the SAS data set containing the item data, the SAS variable names for the items, and the SAS variable name for a case identification variable. Additional arguments indicate whether or not the total scores should be rounded (yes or no), whether the item scores should be combined using a mean or a sum (mean or sum), and whether the data should be printed after the totals have been computed (yes or no). By default, the macro produces a SAS data set named FINAL that includes all of the original data augmented with the four total scores.

```
%macro missing_items (data = _LAST_, vars = x1 - x10, id = idn, round = no, type =
mean, print = yes);

* +-----------------------------------------------------------------------+
    Macro arguments:
            data  = name of SAS data set with rating items
            vars  = SAS variable names of rating items
            id    = case identification variable (numeric)
            round = request rounding of final values (x.xx) -- yes or no
            type  = type of total score to compute -- sum or mean
            print = request printing of final data set created -- yes or no
  +-----------------------------------------------------------------------+;

proc sort data = &data;
      by &id;
data temp;
      set &data;
      keep &vars;

proc mi noprint data=&data out=imputed;
```

```
proc sort data = imputed;
      by &id;
proc means noprint data = imputed;
      by &id;
      output out = mi_means mean = ;
data mi_means2;
      set mi_means;
      keep &vars;
run;

proc iml;
      use &data;
      read all var {&id} into id_vector;
      use temp;
      read all into obsdata;
      use mi_means2;
      read all into mi_data;

      samp = NROW(obsdata);
      items = NCOL(obsdata);

      Impdata = J(samp, 5, 0); * Matrix to hold the total scores;

* +---------------------------------------------+
          Person Mean Substitution
  +---------------------------------------------+;

      do row = 1 to samp;
             PMS = 0;
             Nvalues = 0;
             do column = 1 to items;
                    if (obsdata[row,column] ^= .) then do;
                           PMS = PMS + obsdata[row, column];
                           Nvalues = Nvalues + 1;
                    end;
             end;
                    PMS = PMS/Nvalues;
                    Impdata [row, 1] = PMS;
                    Impdata [row, 2] = PMS;
                    Impdata [row, 3] = PMS;
                    Impdata [row, 5] = items - Nvalues;
      end;
* +---------------------------------------------+
         Single Regression Substitution
  +---------------------------------------------+;
      * +-----------------------------------------------------------------+
         Compute Pairwise Correlations from the non-missing observations in the
         Sample. The correlation coefficients, means, and variances will be saved
         in a matrix called set_of_corrs
         +-----------------------------------------------------------------+;
      do var1 = 1 to items;
             do var2 = 1 to items;
                    if var1 < var2 then do;
                           do row = 1 to samp;
                                  if (obsdata[row,var1] ^= . & obsdata[row,var2]^= .)
                                  then pairdata = pairdata //( obsdata[row,var1] ||
                                  obsdata[row,var2]);
                           end;
* +----------------------------------------------------------------------------
+  Compute means, variances, and correlation based on little set of complete data
   on the pair of variables
   +----------------------------------------------------------------------------+;
                           paircorr1 = corr(pairdata);
                           paircorr = paircorr1[1,2];
```

```
                            n_obs = nrow(pairdata);
                            means = (J(1,n_obs,1)*pairdata)/N_obs;
                            S2 = 1/(n_obs-1)*pairdata`*(i(n_obs)-1/n_obs* j(n_obs))*
                                    pairdata;
                            if (paircorr ^= 0) then set_of_corrs = set_of_corrs //
                                    (var1||var2||paircorr||means||s2[1,1]||s2[2,2]);
                            free pairdata;
                    end;
            end;
        end;
* +-------------------------------------------------------------------------+
        For each missing value:
        1. select the single best predictor among the non-missing items for the
            observation
        2. estimate the sample regression equation using that predictor
        3. use the sample equation to predict the missing value
        4. impute the predicted value
    +-------------------------------------------------------------------------+;
        do row = 1 to samp;
                SRS = 0;
                Nvalues = 0;
                do column = 1 to items;
                        if (obsdata[row,column] ^= .) then do; * No missing data here;
                                SRS = SRS + obsdata[row,column];
                                Nvalues = Nvalues + 1;
                        end;
                        if (obsdata[row,column]  = .) then do; * missing item at this
point in the data;
                                best_r = {0 0 0 0 0 0 0};
                                do search = 1 to nrow(set_of_corrs); * search through
                                                            set_of_corrs for the highest
                                                            correlation with COLUMN;
                                    if set_of_corrs[search,1] = column |
                                      set_of_corrs[search,2] = column then do;
                                            if (obsdata[row,set_of_corrs[search,1]] ^= . |
                                              obsdata[row,set_of_corrs[search,2]] ^= .)
                                              then do;
                                                    if ABS(set_of_corrs[search,3]) >
                                                            ABS(best_r[1,3]) then do;
                                                            best_r = set_of_corrs[search,];
                                                    end;
                                            end;
                                    end;
                                end;
                                if best_r[1,1] = column then do; * The first variable in
                                                        best_r will be the criterion variable;
                                        slope = best_r[1,3]#(SQRT(best_r[1,6])/
                                                SQRT(best_r[1,7]));
                                        intercept = best_r[1,4]-(slope#best_r[1,5]);
                                        Y_hat = intercept + slope#obsdata[row,best_r[1,2]];
                                end;
                                if best_r[1,2] = column then do; * The second variable in
                                                        best_r will be the criterion variable;
                                        slope = best_r[1,3]#(SQRT(best_r[1,7])/
                                                SQRT(best_r[1,6]));
                                        intercept = best_r[1,5]-(slope#best_r[1,4]);
                                        Y_hat = intercept + slope#obsdata[row,best_r[1,1]];
                                end;
                                SRS = SRS + Y_hat;
                        end;
                end;
                if (Nvalues < items) then do; * An imputation was made for this
                                        observation;
```

7

```
                SRS = SRS/items;
                Impdata [row, 2] = SRS;* Place imputed value in the matrix;
            end;
        end;
        free set_of_corrs;
* +-------------------------------------------+
        Relative Mean Substitution
  +-------------------------------------------+;
        Missing1 = 0;
* +------------------------+
        Compute all item sums
  +------------------------+;
        item_sums = J(1,items,0);
        item_ns = J(1,items,0);
        do column = 1 to items;
            do row = 1 to samp;
                if (obsdata[row,column] ^= .) then do;
                    item_sums[1,column] = item_sums[1,column] +
                        obsdata[row,column];
                    item_ns[1,column] = item_ns[1,column] + 1;
                end;
            end;
        end;

        do row = 1 to samp;
            RMS = 0;
            Nvalues = 0;
            * +-------------------------------------------------------------+
                Create lists of observed and missing items for this person
              +-------------------------------------------------------------+;
            missing_items = {0};
            observed_items = {0};
            do column = 1 to items;
                if (obsdata[row,column] ^= .) then do;  * No missing value for
                                                          this item;
                    observed_items = observed_items || column; * List of
                                                          observed items;
                    RMS = RMS + obsdata[row, column];      * Sum of observed
                                                    item values for each person;
                    Nvalues = Nvalues + 1;              * Number of items for each
                                                      person;
                end;
                if (obsdata[row,column] = .) then do;  * Missing value for this
                                                          item;
                    missing_items = missing_items || column; * List of missing
                                                              items;
                end;
            end;
            * +-----------------------------------------------------+
                Compute numerator and denominator of the RMS formula
              +-----------------------------------------------------+;
            if ncol(missing_items) > 1 then do; * at least one missing item for this
                                            record;
                Person_mean = RMS / Nvalues;
                matrix_mean = 0;
                matrix_n = 0;
                do observed_i = 2 to ncol(observed_items);
                    matrix_mean = matrix_mean +
                        item_sums[1,observed_items[1,observed_i]] -
                        obsdata[row,observed_items[1,observed_i]];
                    matrix_n=matrix_n+item_ns[1,observed_items[1,observed_i]]-
1;
                end;
```

```
                  matrix_mean = matrix_mean / matrix_n;
            * +-------------------------------------------+
                  Impute missing items with the RMS formula
              +---------------------------------------------+;
                  do missing_i = 2 to ncol(missing_items);
                          RMS = RMS + (person_mean/matrix_mean)#
                                  (item_sums[1,missing_items[1,missing_i]]/
                                   item_ns[1,missing_items[1,missing_i]]);
                  end;
                  RMS = RMS/items;
                  Missing1 = missing1 + 1;
                  Impdata [row, 3] = RMS;* Place imputed value in the matrix;
            end;
      end;
* +---------------------------------------------+
        PROC MI Results
  +---------------------------------------------+;
      do row = 1 to samp;
            MI = 0;
            do column = 1 to items;
                  MI = MI + mi_data[row,column];
            end;
            Impdata[row,4] = MI/items;
      end;
      score_type = "&type";
      if score_type = 'sum' then do;
            do row = 1 to samp;
                  do col = 1 to 4;
                          impdata[row,col] = impdata[row,col]#items;
                  end;
            end;
      end;
* +---------------------------------------------------+
    Send simulated samples to regular SAS for analysis
  +---------------------------------------------------+;
      outdata = ID_Vector||impdata;
      cname = {"&ID" "PMS" "SRS" "RMS" "MI" "N_Miss"};
      create TEST_FILE from outdata [ colname=cname ];
      append from outdata;
      free outdata;
quit;
data final;
      merge &data test_file;
      by &id;
      %if &round = yes %then %do;
            PMS = ROUND(PMS,.01);
            SRS = ROUND(SRS,.01);
            RMS = ROUND(RMS,.01);
MI = ROUND(MI,.01);
      %end;
      label       PMS = 'Person Mean Substitution'
                  SRS = 'Simple Regression Imputation'
                  RMS = 'Relative Mean Substitution'
                  MI  = 'Multiple Imputation';
%if &print = yes %then %do;
      proc print data = final label;
            title 'This is data set final';
%end;
run;
%mend missing_items;
```

MACRO EXECUTION

As an example of the use of the macro MISSING_ITEMS, the following SAS data step creates a SAS data set called ONE, containing five observations. Each observation has responses to five individual items (X1 – X5), as well as an

9

observation identifier (ID). Missing data have occurred on items X4 and X5 for three of the cases. The macro MISSING_ITEMS is called twice to illustrate the difference between computing total scores as the sum versus the mean. In each call to the macro, the data set is identified (data = one), and both the SAS names for the item variables (vars = X1 – X5) and the case identifier variable (id = id) are provided. In each macro call, rounding is requested (round = yes) and a printed list of cases is requested (print = yes). The only difference in the two macro calls is the method by which total scores will be produced (type = mean in the first macro call, and type = sum in the second).

```
data one;
 input ID x1 x2 x3 x4 x5;
cards;
1  1 2 3 4 5
2  2 2 3 . 5
3  3 1 4 2 .
4  3 3 2 . .
5  2 2 1 4 3
;
%missing_items (data=one, vars=x1 - x5, id=id, round = yes, type = mean, print = yes);
%missing_items (data=one, vars=x1 - x5, id=id, round = yes, type = sum, print = yes);
run;
```

OUTPUT EXAMPLES OF MACRO MISSING_ITEMS

The macro is called twice to illustrate the difference between computing person total score as means versus computing person total score as sums. Output 1, as requested, is printed to show the estimates of person total scores as means, obtained using multiple imputation (*MI*), single regression substitution (*SRS*), relative mean substitution (*RMS*), and person mean substitution (*PMS*) missing data methods. The number of missing items for each observation is also reported. Note that the original data have not been altered – items which were missing in the data set are still missing. The macro has added to the original data set the new variables to represent total scores for each observation. Because the 'round = yes' option was used with the macro, the total scores are rounded to the nearest hundredth. Table 5 below shows sample output from first macro call.

```
                              This is data set final

                                         Simple      Relative
                              Person Mean Regression    Mean       Multiple
Obs  ID  x1  x2  x3  x4  x5   Substitution Imputation Substitution Imputation  N_Miss
1    1   1   2   3   4   5       3.00        3.0        3.00         3.00         0
2    2   2   2   3   .   5       3.00        3.2        3.00         3.07         1
3    3   3   1   4   2   .       2.50        3.2        2.50         2.85         1
4    4   3   3   2   .   .       2.67        3.6        2.67         3.12         2
5    5   2   2   1   4   3       2.40        2.4        2.40         2.40         0
```
Output 1. Example output from macro when computing total scores as means

Output 2, as requested, is printed to show the estimates of person total scores as sums obtained using multiple imputation (*MI*), single regression substitution (*SRS*), relative mean substitution (*RMS*), and person mean substitution (*PMS*) missing data methods. As with the first macro output, the total scores in the output have been rounded to the nearest hundredth.

```
                              This is data set final

                                         Simple      Relative
                              Person Mean Regression    Mean       Multiple
Obs  ID  x1  x2  x3  x4  x5   Substitution Imputation Substitution Imputation  N_Miss
1    1   1   2   3   4   5      15.00        15        15.00        15.00         0
2    2   2   2   3   .   5      15.00        16        15.00        15.33         1
3    3   3   1   4   2   .      12.50        16        12.50        14.33         1
4    4   3   3   2   .   .      13.33        18        13.33        15.67         2
5    5   2   2   1   4   3      12.00        12        12.00        12.00         0
```
Output 2. Example output from macro when computing total scores as sums

CONCLUSION

Raaijmakers (1999) denotes missing data as a common problem in empirical research. Although the development of methods for the prevention and minimization of nonresponse rates in early steps of the research process (e.g., design, sample selection process, and data collection), it is unlikely that all data will be available for all persons (unit nonresponse) and or for all items (item nonresponse). The effect of nonresponse at the person and item levels on point estimators is of concern. For instance, means and sums can be biased. Also, the variance of these estimators can be under or overstated. Consequently, the implementation of adequate methods for handling missing data is necessary for obtaining correct statistical inferences.

While there is a vast literature on missing data methods, unfortunately, researchers seem to be implementing simpler yet less efficient missing data treatments (e.g., Listwise and Pairwise deletion methods), maybe in part because these are the default methods offered by popular statistical software packages (O'Rourke, 2003; Roth & Switzer, 1995). At the same time, as Rubin (2003) states, "when confronted with missing data, it is a hopelessly daunting task to derive new methods of data analysis" (p. 619). Thus, the primary purpose of this paper was to offer to interested users an easy implementation of four missing data methods for Likert-type scales. The prospect of using all information available in their data sets by not having to use deletion methods should be encouraging.

REFERENCES

Allison, P. D. (2002). *Missing Data.* Thousand Oaks, CA: Sage Publications.

Beale, E. M. L, & Little, R. J. A. (1975). Missing values in multivariate analysis. *Journal of the Royal Statistical Society*, Series B (Methodological), *37*(1), pp. 129-145.

Brockmeier, L. L., Kromrey, J. D., & Hogarty, K. Y. (2003). Nonrandomly missing data in multiple regression analysis: An empirical comparison of ten missing data treatments. *Multiple Linear Regression Viewpoints, 29*(1), 8-29.

Buhi, E. R., Goodson, P., Neilands, T. B. (2008). Out of sight, not out of mind: Strategies for handling missing data. *American Journal of Health Behavior, 32*(1), 83-92.

Croker, L., & Algina, J. (1986). *Introduction to Classical & Modern Test Theory*. Belmont: CA, Thompson Learning.

Downey, R. G., & King, C. V. (1998). Missing data in likert ratings: A comparison of replacement methods. *The Journal of General Psychology, 125*(2), 175-191.

Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician, 61*(1), 79-90.

Kim, J., & Curry, J. (1977). The treatment of missing data in multivariate analysis. *Sociological Methods and Research, 22,* 896.

Kromrey, J. D., & Hines, C. V. (1994). Nonrandomly missing data in multiple regression: An empirical comparison of common missing-data treatments. *Educational and Psychological Measurement, 54*(3), 573-593.

Little, R. J. A. (1978). Consistent regression methods for discriminant analysis with incomplete data. *Journal of the American Statistical Association, 73* (362), pp. 319-322.

O'Rourke, T. W. (2003). Methodological techniques for dealing with missing data. *American Journal of Health Studies, 18*(2/3), 165-168.

Patrician, P. A. (2002). Multiple imputation for missing data. *Research in Nursing & Health, 25*, 76-84.

Raaijmakers, Q. A. W. (1999). Effectiveness of different missing data treatments in surveys with likert-type data: Introducing the relative means substitution approach. *Educational and Psychological Measurement*, *59*(5), 725-748.

Rodriguez, P. & Kromrey, J. D. (2010). An empirical comparison of four missing data techniques for estimating person scores on Likert-type scales. Paper presented at the 55th Annual Conference of the Florida Educational Research Association, Orlando, FL.

Roth, P. L., & Switzer, F. S. (1995). A Monte Carlo analysis of missing data techniques in a HRM setting. *Journal of Management, 21*(5), 1003-1023.

Rubin, D. B. (2003). Discussion on multiple imputation. *International Statistical Review, 71*(3), 619-625.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*, 3-15.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*(4), 545-571.

Shrive, F.M., Stuart, H., Quan, H., & Ghali, W. A. (2006). Dealing with missing data in a multi-question depression scale: A comparison of imputation methods. *BMC Medical Research Methodology, 6, 1-10.*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the first author at:

Patricia Rodríguez de Gil   Work Phone: (813) 974-3220
University of South Florida   Fax: (813) 974-4495
4202 East Fowler Ave., EDU 105   Email: prodrig6@usf.edu
Tampa, FL 33620