

Using SAS ® to Examine Missing Data in Psychometric Research

Abbas S. Tavakoli, DrPH, MPH, Elizabeth Register, PhD, MPH, RN-BC, CCM, & JoAnne Herman, PhD, RN, CSME

Abstract

Background:

This study examined the effect of missing data in psychometric research. Missing data present a challenge to researchers. Incomplete data is common in most researches. There are many reasons for no response such as poor research designs, poor structured questions, and attrition in longitudinal studies.

Method:

A convenience sample of 428 community dwelling older adults participated in this study. Scale items were generated through a grounded theory study that was conducted to explore connectedness in older adults. The Register - Connectedness Scale for Older Adults is a 72-likert item self-administered paper and pencil survey. Single and multiple imputations were used to replace the missing data for 72 items. Means of each item were compared with and without imputation. Three factor analyses were run to develop Register -Connectedness Scales for older adults. These run included no imputation, single imputation, and multiple imputation for missing data.

Results:

The missing data for the 72 items ranges from 4-34. The results indicated there were no difference between means of items no imputation, single imputation, and multiple imputations. Optimal factor solution involved a five-factor structure which yielded a 55-item scale. Factor analyses results indicated all items were loaded exactly the same for single imputation and multiple imputations. However, there was a difference in the loading of items for four of five factors when no imputation was compared with imputation methods. Some of items did not load in the without imputation where they loaded in the imputations methods. Also, few items loaded in without imputation in different factors compare to with imputations methods.

Conclusion:

Our study indicated that there were differences in the result of factor analysis when imputation used for replacing missing data compare to no imputation. Researchers should consider using imputation method to help improve problems caused by missing data in the study.

Keywords:

Missing Data, Imputation, Factor Analysis.

Introduction

Most SAS/STAT® Statistical procedures eliminate the data point with any missing value from the analysis. Observations with missing data for one or more variables called incomplete cases. Using incomplete cases in statistical analysis may not be applicable to the population of all cases. Missing data in the data present a challenge to researchers in the data analysis phase. Incomplete data is common in most researches. There are many reasons for no response such as poor research designs, poor structured questions, and attrition in longitudinal studies.

Purpose

This study presentation examines the effect of the missing data in psychometric research.

Background

A convenience sample of 428 community dwelling older adults (65-98 years) participated in this study. Participants were recruited from senior/community centers and congregate meal sites churches, retirement communities, formal groups, informal groups, and public housing located in the Southeastern states of South Carolina, Georgia, and Florida. Instrument items were generated through a grounded theory study that was conducted to explore connectedness in older adults. The Register - Connectedness Scale for Older Adults is a 72-item self-administered paper and pencil survey. The items were short declarative statements. Design features were included to increase readability and ease of use because the instrument used for older adults. Data were collected both individually and in groups. At every site, the researcher gave an overview of the study to ensure participants met the inclusion criteria and understood that completing the instrument was giving informed consent. The inclusion criteria was older adults had to be community dwelling, age 65 or older, and English speaking or reading.

Data Analyses

All data analyses were performed using SAS/STAT® statistical software, version 9.2 (SAS, 2008). Single and multiple imputations were used to replace the missing value for 72 items. Means of each item was compared with and without imputation. Three factor analyses were run to develop Register -Connectedness Scales for older adults. These run included no imputation, single imputation, and multiple imputation for missing data. Factor analysis used squared multiple correlations as prior communality estimates. The maximum Likelihood (ML) method was used to extract the factors. And this was followed by the promax (oblique) rotation. In interpreting the rotated factor pattern, an item was said to load on a given factor if the factor loading was .35 or greater for that factor, and was less than .35 for the other.

Results

The frequency distribution indicated that the missing value for 72 items were range from 4-34. The results indicate there was no difference between means of no imputation, single imputation, and multiple imputation (see table 1). A scree test, Eigen values, and the proportion variance explained with each factor suggested five meaningful factors which yielded a 55-item scale. Factor analysis results indicate all items were loaded exactly the same for single imputation and multiple imputations. However, there was a difference in the loading of items for four of five factors when no imputation was compared with imputation methods. Some of items did not load in the without imputation where they loaded in the imputations methods (items 16, 53, 65, 69, and 72). Also, few items loaded in without imputation in different factors compare to with

imputation methods (item 15 loaded in factor 3, item 27 and 36 loaded in factor2, and item 70 loaded in factor 4) (see table 2).

Conclusion

Our study indicates that there was difference in the result of factor analysis when imputation used for replacing missing data compare to no imputation. There are several methods for replacing missing value in the analysis. Single imputation method is due to bias if the proportion of missing value is large (greater than 5%). Multiple imputations (MI) are another alternative method to replace missing value. Multiple imputation method is more accurately reflect the uncertainly due to missing data. Researchers should consider using imputation method to help improve problems caused by missing data in the study.

Table1. N, mean, and Standard deviation for selected items with imputation, and no imputation

Items	Description of Items	Imputation (N=428)		No Imputation		
		Mean	STD	N	Mean	STD
RCSI1	Spent time with other people	3.561	0.728	424	3.561	0.732
RCSI2	Thought about the old times	2.940	0.972	415	2.940	0.987
RCSI3	Given or received a hug or kiss	3.401	0.854	414	3.401	0.868
RCSI4	Had a reason to get out of bed	3.703	0.673	417	3.703	0.681
RCSI5	Helped someone	3.597	0.679	419	3.597	0.686
RCSI6	Wanted to be around others	3.498	0.716	420	3.498	0.723
RCSI7	Wanted to talk to someone	3.390	0.770	418	3.390	0.780
RCSI8	Felt I am part of the world	3.553	0.744	423	3.553	0.749
RCSI9	Known my life has a purpose	3.689	0.686	424	3.689	0.689
RCSI10	Seen the good in myself	3.445	0.794	418	3.445	0.804
RCSI11	Wished I were more productive	3.207	0.900	415	3.207	0.914
RCSI12	Thought about getting old	2.733	1.103	416	2.733	1.119
RCSI13	Felt needed by someone	3.437	0.826	419	3.437	0.834
RCSI14	Felt wanted by someone	3.449	0.827	421	3.449	0.834
RCSI15	Felt loved by someone important to me	3.755	0.558	417	3.755	0.566
RCSI16	Seen the good in someone	3.655	0.589	417	3.655	0.597
RCSI17	Felt appreciated by others	3.518	0.733	423	3.518	0.738
RCSI18	Felt that my life has been worthwhile	3.739	0.563	422	3.739	0.567
RCSI19	Been part of a religious group	3.596	0.772	418	3.596	0.781
RCSI20	Believed that God blesses me	3.829	0.519	414	3.829	0.527
RCSI21	Felt a spiritual connection in nature	3.557	0.738	415	3.557	0.749

Items	Description of Items	Imputation (N=428)		No Imputation		
		Mean	STD	N	Mean	STD
RCSI22	Spent quiet time with myself	3.532	0.721	417	3.532	0.730
RCSI23	Been aware of a higher power	3.733	0.648	415	3.733	0.658
RCSI24	Found comfort in my faith	3.749	0.627	415	3.749	0.637
RCSI25	Thought about people who I care about that have died	3.327	0.887	419	3.327	0.897
RCSI26	Prayed for something important to me	3.659	0.714	416	3.659	0.724
RCSI27	Thought about how my death will affect other people	3.036	1.006	415	3.036	1.022
RCSI28	Displayed family photos in my home	3.474	0.875	422	3.474	0.882
RCSI29	Spent time with my family	3.627	0.704	418	3.627	0.713
RCSI30	Wanted to be with my family	3.580	0.731	419	3.580	0.739
RCSI31	Talked with a family member	3.745	0.547	420	3.745	0.552
RCSI32	Felt close to a family member who lives far away	3.521	0.816	411	3.521	0.833
RCSI33	Felt loved by my family	3.764	0.570	416	3.764	0.579
RCSI34	Wished my family lived closer to me	3.242	0.990	418	3.242	1.002
RCSI35	Felt needed by my family	3.501	0.769	417	3.501	0.779
RCSI36	Worried about a family member	3.353	0.884	419	3.353	0.894

Table2. Items difference for factor analysis for no imputation and imputation (Multiple/single).

Factors	No Imputation	Imputation (multiple/single)
Factor 1	1,3-10, 46, 67, 71	1,3-10, 15, 16, 46, 53, 65, 67, 70,71, 72
Factor 2	27, 36, 51, 52, 55-63, 69	27, 5,1, 52, 55-63, 69
Factor 3	13, 14, 15, 17, 28, 29, 30-36	13, 14, 17, 28, 29, 30-36
Factor 4	37,39,40-44,70	37,39,40-44
Factor 5	19, 20, 23, 24, 26	19, 20, 23, 24, 26

Note: highlight is the differences for item loading with and without imputation.

SAS Syntax

```
** mean of items without imputation ***;  
ods rtf;  
ods listing close;  
proc means data=two n mean std noprint maxdec=3;  
  var rcsi1-rcsi72 ;  
  title ' means/ ' ;  
  title2 'Scale for Older Adults';  
output out=meantest (drop=_type_ _freq_);  
run;  
  
proc transpose data=meantest  
  out=meantest (rename=(col1=n col2=min col3=max col4=mean col5=std));  
run;  
  
proc print data=meantest noobs ;  
  var _name_ _label_ n mean std ;  
format mean 6.3 std 6.3;  
run;  
ods rtf close;  
ods listing;  
quit;  
run;  
**** single Imputation ****;  
ods rtf;  
ods listing close;  
  
Proc STANDARD DATA=two OUT=stnd REPLACE PRINT ;  
  VAR rcsi1-rcsi72;  
  title ' standard means/replacement of missing ' ;  
  title2 'Scale for Older Adults';  
Run;  
proc means data=stnd n mean std noprint maxdec=3;  
  var rcsi1-rcsi72 ;  
  title ' replacement of missing by means ' ;  
  title2 'Scale for Older Adults';  
  output out=meantest (drop=_type_ _freq_);  
run;  
proc transpose data=meantest  
  out=meantest (rename=(col1=n col2=min col3=max col4=mean col5=std));  
format mean 6.3 std 6.3;  
run;  
  
proc print data=meantest noobs ;
```

```

    var _name_ _label_ n mean std ;

run;
ods rtf close;
ods listing;
quit;
run;

**** Multiple imputation ****;
proc mi data=two seed=37851 out=outmi noprint;
var rcsi1-rcsi72;
title ' Multiple imputation ';
    title2 'Scale for Older Adults';
run;
proc univariate data=outmi noprint;
var rcsi1-rcsi72;
output out=outuni mean =mrcsi1-mrcsi72;
run;
data all;
if _N_ = 1 then set outuni(keep = mrcsi1-mrcsi72);
set two;
data final;
set all;
array items rcsi1-rcsi72;
    array itemsb mrcsi1-mrcsi72;
    do over items;
        if items =. then items=itemsb;
    end;

ods rtf; ods listing close;
proc means data=final n mean std noprint maxdec=3;
var rcsi1-rcsi72 ;
title ' means/mi imputation ';
title2 'Scale for Older Adults';
output out=meantest (drop=_type_ _freq_);
run;
proc transpose data=meantest
out=meantest (rename=(col1=n col2=min col3=max col4=mean col5=std));
format mean 6.3 std 6.3;
run;
proc print data=meantest noobs ;
var _name_ _label_ n mean std ;
run;
ods rtf close; ods listing; quit; run;

```

```
*** SYntax for factor analysis ***;
ods rtf; ods listing close;
title;
%macro fact (d,q,n,t);
proc factor data=&d method=ml priors=smc scree rotate=promax reorder
    flag=.35 nfact=&n ;
    var &q ;
    title ' factor analysis/priors=smc ' &t;
    title2 'Scale for Older Adults/ ';

%mend fact;
%fact (final,rcsi1-rcsi72,5, how important/mi imputation);
%fact (stnd,rcsi1-rcsi72,5, how important/single imputation);
%fact (two,rcsi1-rcsi72,5, how important/no imputation);
run;
ods rtf close; ods listing; quit; run;
PART OF SAS OUTPUT
```

Rotated Factor Pattern (Standardized Regression Coefficients)/No Imputation

		Factor1	Factor2	Factor3	Factor4	Factor5
RCSI57	Thought the world is changing	74 *	-15	10	-13	10
RCSI60	Thought about what a different world it is today	68 *	-5	-3	-4	9
RCSI58	Followed a special news story	66 *	9	5	-3	-8
RCSI63	Talked about current events with someone	63 *	-8	7	11	0
RCSI61	Worried about our country's future	63 *	-7	2	-3	8
RCSI62	Worried about our children's future	54 *	6	1	-11	14
RCSI52	Thought about my health	51 *	8	-1	-3	3
RCSI56	Watched the news on TV	49 *	2	12	-8	-3
RCSI59	Thought about people in need	47 *	7	14	11	5
RCSI36	Worried about a family member	42 *	31	-23	18	-16
RCSI27	Thought about how my death will affect other people	42 *	32	-7	9	-2
RCSI45	Thought about a friend who has recently died	40 *	6	-13	37 *	-9
RCSI51	Wished my health were better	39 *	13	-4	1	8
RCSI55	Kept up with current events	39 *	-1	8	7	1
RCSI12	Thought about getting old	39 *	14	-2	10	-9
RCSI25	Thought about people who I care about that have died	37 *	34	-25	9	5
RCSI66	Talked on the phone to someone	35 *	9	8	9	-3
RCSI50	Felt good about my doctor	33	-1	27	-4	8
RCSI69	Wanted to go somewhere	32	-10	12	27	1
RCSI47	Read food labels	30	4	4	9	-1
RCSI49	Drank plenty of water	25	-6	16	13	0
RCSI68	Enjoyed spending time with a pet	23	-5	-17	20	0
RCSI22	Spent quiet time with myself	23	6	14	10	21
RCSI2	Thought about the old times	22	7	10	16	-9
RCSI29	Spent time with my family	-4	79 *	-5	-1	9
RCSI30	Wanted to be with my family	1	79 *	-15	0	5
RCSI33	Felt loved by my family	-6	71 *	-2	4	2
RCSI31	Talked with a family member	-1	69 *	3	-5	3
RCSI35	Felt needed by my family	8	69 *	7	-3	1
RCSI13	Felt needed by someone	0	59 *	18	4	-8
RCSI14	Felt wanted by someone	-9	55 *	21	7	-10
RCSI32	Felt close to a family member who lives far away	20	50 *	4	-1	-1
RCSI15	Felt loved by someone important to me	0	49 *	17	-12	9
RCSI17	Felt appreciated by others	2	46 *	24	7	4
RCSI34	Wished my family lived closer to me	30	41 *	-1	15	-3
RCSI18	Felt that my life has been worthwhile	0	38 *	22	6	20
RCSI28	Displayed family photos in my home	24	35	4	0	-10

Rotated Factor Pattern (Standardized Regression Coefficients)/No Imputation

		Factor1	Factor2	Factor3	Factor4	Factor5	
RCSI54	Felt rested after sleeping	4	35	6	-3	16	
RCSI11	Wished I were more productive	13	26	19	8	-8	
RCSI64	Driven a car	7	-11	-3	5	10	
RCSI7	Wanted to talk to someone	-3	-4	79 *	5	-5	
RCSI6	Wanted to be around others	-3	7	74 *	-5	-4	
RCSI8	Felt I am part of the world	4	-4	70 *	4	2	
RCSI1	Spent time with other people	-21	8	58 *	12	-6	
RCSI5	Helped someone	-8	32	53 *	-4	3	
RCSI9	Known my life has a purpose	-2	18	52 *	-13	23	
RCSI10	Seen the good in myself	25	19	44 *	6	-11	
RCSI67	Taken a walk indoors or outdoors	23	-13	44 *	12	-11	
RCSI3	Given or received a hug or kiss	-6	12	43 *	15	1	
RCSI71	Spent time doing something I enjoy	20	-4	37 *	12	6	
RCSI46	Exercised with others or alone	8	-3	37 *	-1	-1	
RCSI65	Listened to music or sang a song	17	-15	32	12	17	
RCSI4	Had a reason to get out of bed	14	21	32	-14	-3	
RCSI16	Seen the good in someone	8	23	30	4	21	
RCSI72	Made my own decisions	7	1	15	3	7	
RCSI48	Taken my medicine on time	3	8	12	3	11	
RCSI41	Felt good about having friends close by	-9	-1	2	84 *	13	
RCSI42	Wanted to be with friends	0	-5	6	82 *	1	
RCSI43	Felt good about my friendships	3	-2	-7	80 *	8	
RCSI40	Spent time with a friend	-7	-1	11	77 *	5	
RCSI39	Thought about a friend	3	7	12	59 *	4	
RCSI37	Talked with an old friend	-1	7	10	57 *	5	
RCSI44	Wished I had more friends close by	25	5	-2	47 *	-4	
RCSI70	Found humor in something	16	-5	18	40 *	-8	
RCSI38	Sent or received a letter or an e-mail	6	19	3	33	-9	
RCSI24	Found comfort in my faith	8	-1	-11	1	90 *	
RCSI23	Been aware of a higher power	6	-5	6	-5	82 *	
RCSI20	Believed that God blesses me	3	-2	1	2	77 *	
RCSI26	Prayed for something important to me	2	3	6	7	71 *	
RCSI19	Been part of a religious group	-22	22	-10	11	57 *	
RCSI21	Felt a spiritual connection in nature or outdoors	20	4	27	4	29	
RCSI53	Taken care of myself	6	9	22	-4	28	

Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.35 are flagged by an '*'. Values greater than 0.70 are flagged by a '**'.

Rotated Factor Pattern (Standardized Regression Coefficients) / Imputation

		Factor1	Factor2	Factor3	Factor4	Factor5
RCSI8	Felt I am part of the world	59 *	-2	7	2	5
RCSI9	Known my life has a purpose	56 *	-3	15	-4	13
RCSI71	Spent time doing something I enjoy	54 *	18	-6	9	-5
RCSI5	Helped someone	54 *	-8	18	0	5
RCSI6	Wanted to be around others	50 *	-3	18	3	-2
RCSI7	Wanted to talk to someone	47 *	7	10	8	3
RCSI67	Taken a walk indoors or outdoors	47 *	15	-13	8	-7
RCSI72	Made my own decisions	45 *	5	-1	5	-2
RCSI53	Taken care of myself	44 *	8	1	1	11
RCSI65	Listened to music or sang a song	42 *	11	-11	12	8
RCSI16	Seen the good in someone	40 *	7	14	3	8
RCSI1	Spent time with other people	40 *	-26	10	13	9
RCSI10	Seen the good in myself	40 *	18	23	2	-3
RCSI15	Felt loved by someone important to me	39 *	4	39 *	-18	-4
RCSI4	Had a reason to get out of bed	39 *	3	19	-11	0
RCSI3	Given or received a hug or kiss	39 *	-5	18	3	4
RCSI46	Exercised with others or alone	38 *	-1	-3	6	1
RCSI70	Found humor in something	36 *	14	-8	25	-7
RCSI54	Felt rested after sleeping	33	-7	19	6	5
RCSI18	Felt that my life has been worthwhile	32	5	30	3	2
RCSI50	Felt good about my doctor	31	19	5	5	3
RCSI49	Drank plenty of water	27	15	0	4	-2
RCSI64	Driven a car	25	11	-8	-8	-1
RCSI60	Thought about what a different world it is today	-1	66 *	-1	-1	12
RCSI57	Thought the world is changing	13	63 *	-9	-5	5
RCSI58	Followed a special news story	15	63 *	-1	0	-5
RCSI61	Worried about our country's future	9	60 *	3	-10	1
RCSI63	Talked about current events with someone	27	56 *	-7	6	-8
RCSI62	Worried about our children's future	15	55 *	5	-17	9
RCSI52	Thought about my health	1	46 *	20	-4	4
RCSI51	Wished my health were better	-8	43 *	29	-4	2
RCSI56	Watched the news on TV	16	42 *	0	-4	-3
RCSI59	Thought about people in need	17	41 *	2	10	15
RCSI27	Thought about how my death will affect other people	-16	38 *	38 *	7	1
RCSI55	Kept up with current events	31	36 *	-10	3	-11
RCSI69	Wanted to go somewhere	8	35 *	-3	16	-3
RCSI45	Thought about a friend who has recently died	-16	33	22	32	-7

Rotated Factor Pattern (Standardized Regression Coefficients) / Imputation

		Factor1	Factor2	Factor3	Factor4	Factor5
RCSI25	Thought about people who I care about that have died	-23	33	25	16	15
RCSI66	Talked on the phone to someone	22	29	1	8	-3
RCSI12	Thought about getting old	-8	29	28	4	-9
RCSI47	Read food labels	17	23	7	3	1
RCSI68	Enjoyed spending time with a pet	-15	20	2	14	5
RCSI2	Thought about the old times	-1	18	19	14	-1
RCSI30	Wanted to be with my family	-4	-3	76 *	-2	6
RCSI35	Felt needed by my family	-1	5	67 *	5	5
RCSI33	Felt loved by my family	8	-7	64 *	1	9
RCSI29	Spent time with my family	23	-9	61 *	-7	4
RCSI31	Talked with a family member	22	-7	56 *	-5	7
RCSI13	Felt needed by someone	30	-1	52 *	2	-18
RCSI34	Wished my family lived closer to me	-10	25	49 *	13	2
RCSI14	Felt wanted by someone	30	2	47 *	3	-20
RCSI32	Felt close to a family member who lives far away	16	12	42 *	4	0
RCSI36	Worried about a family member	-15	25	41 *	17	-8
RCSI28	Displayed family photos in my home	11	11	38 *	0	1
RCSI17	Felt appreciated by others	26	3	35	3	10
RCSI11	Wished I were more productive	1	12	32	11	0
RCSI41	Felt good about having friends close by	13	-11	-6	81 *	5
RCSI42	Wanted to be with friends	2	-1	3	79 *	0
RCSI40	Spent time with a friend	18	-7	0	71 *	-3
RCSI43	Felt good about my friendships	8	-5	1	69 *	12
RCSI39	Thought about a friend	16	10	2	55 *	-4
RCSI37	Talked with an old friend	15	2	1	54 *	2
RCSI44	Wished I had more friends close by	-24	26	21	45 *	0
RCSI38	Sent or received a letter or an e-mail	10	10	6	31	-8
RCSI48	Taken my medicine on time	8	-2	10	14	9
RCSI24	Found comfort in my faith	1	8	-6	-2	86 *
RCSI26	Prayed for something important to me	5	4	2	0	79 *
RCSI20	Believed that God blesses me	-6	-2	9	1	78 *
RCSI23	Been aware of a higher power	10	7	-18	2	77 *
RCSI19	Been part of a religious group	0	-14	17	2	58 *
RCSI22	Spent quiet time with myself	10	17	7	4	28
RCSI21	Felt a spiritual connection in nature or outdoors	19	22	11	-2	27

Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.35 are flagged by an '*'. Values greater than 0.70 are flagged by a '**'.

References

- Allison, P.D. (2000). Multiple Imputation for Missing Data: A Cautionary Tale. *Sociological Methods and Research*, 28,301-309.
- Hatcher, Larry 2003. *Step by Step Approach to Using the SAS System for Factor Analysis* (1994). SAS Inc.
- Horton, N.J, & Lipsitz, S. R. (2001). Multiple Imputation in Practice: Comparison of Software Packages for Regression Models with Missing Variables. *Journal of the American Statistical Association*, 55, 244-254
- Little, R. & Rubin, D. (1989). *Statistical analysis with missing data*. New York: Wiley.

Contact Information

Abbas S. Tavakoli
University of South Carolina College of Nursing
College of Nursing University of South Carolina
1601 Greene Street
Columbia, SC 29208-4001
Fax: (803) 777-5561
E-mail: abbas.tavakoli@sc.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.