IMPUTE Example #2

SUDAAN Statements and Results Illustrated

- WSHD Multiple Imputations
- NOTSORTED option
- IMPBY
- IMPVAR (MULTIMP option)
- IMPNAME
- PRINT and OUTPUT

Input Data Set(s): WIC.SAS7bdat

Example

Using data from the Women, Infants and Children (WIC) study, perform multiple imputations (impute 4 times) for missing values of the length of the mother's hospital stay. Use information on the mother's race and smoking behavior as imputation class data.

Solution

The primary purpose of this example is to illustrate how to perform multiple imputations in a single call to IMPUTE. The SAS-callable version of SUDAAN was used to generate the output in this example.

The following variables from the WIC dataset are of interest in this example:

Variable	Definition
ID	Observation number
ANALWGT1	Final Weight
MOMRACE	Mother's Race (2 levels)
MOMSMK	Mother's Smoking Behavior (2 levels)
MOMHOSP	Length of Mother's Hospital Stay (in days)

Prior to performing WSHD imputation, PROC CROSSTAB is used to review the distribution of the imputation variable, MOMHOSP, and PROC RECORDS is used to print the nonrespondents. The CROSSTAB and RECORDS code and results are shown next.

Exhibit 1. CROSSTAB and RECORDS Code

```
LIBNAME in "\\rtints29\sudaan\data\wicwage";
libname out "c:\11winbetatest\examples";

data one; set in.wicwage;

PROC CROSSTAB data=one design=srs;
  class momrace momsmk momhosp / include=missing;
  tables momrace momsmk momhosp;
  print nsum colper / nsumfmt=f6.0 colperfmt=f7.2 style=nchs;
  rtitle "Distributions";

PROC RECORDS DATA=one;
  subpopx momhosp=.;
  sortby momrace momsmk ;
  setenv colwidth=9 labwidth=10 decwidth=0;
  PRINT id momrace momsmk momhosp;
  rtitle "List the MOMHOSP Non-Respondents";
```

Since we are just interested in sample distributions here, we can specify the SRS design in CROSSTAB. And RECORDS does not use design information at all.

Exhibit 2. First Page of CROSSTAB Results

```
                              S U D A A N
           Software for the Statistical Analysis of Correlated Data
         Copyright      Research Triangle Institute         June 2012
                               Release 11.0.0


DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method,
Assuming a
Simple Random Sample (SRS) Design


Number of observations read    :    953
Denominator degrees of freedom :    952
```

Exhibit 3. CROSSTAB Results: Sample Distribution of Mother's Race

```
Variance Estimation Method: Taylor Series (SRS)

Distributions

by: Mother Race.
---------------------------------------
Mother Race
                       Sample   Col
                       Size     Percent
---------------------------------------
Total                     953    100.00
1                         480     50.37
2                         473     49.63
---------------------------------------
```

Exhibit 4. CROSSTAB Results: Sample Distribution of Mother's Smoking Behavior

```
Variance Estimation Method: Taylor Series (SRS)

Distributions

by: Smoking behavior.
--------------------------------------
Smoking behavior
                        Sample   Col
                        Size     Percent
--------------------------------------
Total                      953    100.00
Missing                     21      2.20
1                          230     24.13
2                          702     73.66
--------------------------------------
```

There are 21 missing values for the imputation class variable mother's smoking.

Exhibit 5. CROSSTAB Results: Sample Distribution of Length of Hospital Stay (Days)

```
Variance Estimation Method: Taylor Series (SRS)

Distributions

by: Length of hospital stay -- mother.
--------------------------------------
Length of hospital
  stay -- mother        Sample   Col
                        Size     Percent
--------------------------------------
Total                      953    100.00
Missing                     21      2.20
0                           25      2.62
1                          320     33.58
2                          350     36.73
3                          142     14.90
4                           48      5.04
5                           22      2.31
6                            7      0.73
7                            4      0.42
8                            2      0.21
10                           1      0.10
11                           1      0.10
12                           3      0.31
13                           7      0.73
--------------------------------------
```

There are 21 missing values for the imputation variable length of hospital stay.

Exhibit 6. RECORDS Results: Length of Hospital Stay Nonrespondents

```
                                S U D A A N
           Software for the Statistical Analysis of Correlated Data
          Copyright      Research Triangle Institute         June 2012
                               Release 11.0.0

For Subpopulation: MOMHOSP = .

List the MOMHOSP Non-Respondents

by: Observation Number.
```

Observation Number	ID	Mother Race	Smoking behavior	Length of hospital stay -- mother
60	34129	1	1	.
63	21127	1	1	.
212	11013	1	1	.
234	41131	1	1	.
242	3127	1	2	.
464	3122	1	2	.
465	21126	1	2	.
468	30124	1	2	.
493	34131	1	2	.
496	22130	1	2	.
502	22136	1	2	.
518	23122	1	2	.
630	21130	1	2	.
631	30123	1	2	.
673	41129	1	2	.
674	41128	1	2	.
767	28125	2	1	.
769	28126	2	2	.
932	11125	2	2	.
933	10017	2	2	.
935	22127	2	2	.

In the next step of this example, we impute the missing values via PROC IMPUTE. The imputation will be performed 4 times. The code is displayed next.

Exhibit 7. IMPUTE Code to get Multiple Imputations

```
PROC IMPUTE data=one seed=6537181 method=wshd icnum notsorted;
  weight analwgt1;
  impby momrace momsmk ;
  impvar momhosp / multimp=4;
  impname momhosp="mhosp";
  impid id;
  idvar / all;
  setenv labwidth=12;
  print / donorstat=default means=default donorsumfmt=f5.0 rsumfmt=f5.0
          absdiffmeanfmt=f7.4 reldiffmeanfmt=f7.4;
  output / impute=default filename=out.multimp filetype=sas replace;
  rtitle "Hot-Deck Multiple Imputations for MOMHOSP";

PROC RECORDS data=out.multimp contents;
  setenv labwidth=10 colwidth=3 decwidth=0 colspce=0;
  print id momrace momsmk donorid1 donorid2 donorid3 donorid4
        momhosp mhosp_1 mhosp_2 mhosp_3 mhosp_4 /
        idfmt=f5.0 donorid1fmt=f5.0 donorid2fmt=f5.0 donorid3fmt=f5.0
        donorid4fmt=f5.0;
  rlabel donorid1="dnr1";
  rlabel donorid2="dnr2";
  rlabel donorid3="dnr3";
  rlabel donorid4="dnr4";
  rlabel momhosp="momhos";
  rlabel momrace="momrac";
  rlabel momsmk="momsmk";
  rlabel mhosp_1="hsp1";
  rlabel mhosp_2="hsp2";
  rlabel mhosp_3="hsp3";
  rlabel mhosp_4="hsp4";
  rtitle "WSHD Multiply Imputed Dataset";

PROC FREQ data=out.multimp;
  tables momhosp*(mhosp_1 mhosp_2 mhosp_3 mhosp_4) mhosp_1 mhosp_2 mhosp_3 mhosp_4 /
        list missing;
  title "WSHD Multiple Imputations:  See How MOMHOSP is Imputation Revised";
```

The NOTSORTED option is used on the PROC statement because the file was not sorted by the IMPBY statement variables.

The IMPVAR statement includes the MULTIMP=4 option. This tells SUDAAN to perform the WSHD imputation procedure 4 times. The output dataset will include the results of all 4 of the multiple imputations.

The IMPBY statement specifies MOMRACE and MOMSMK as the imputation class variables.

The IMPNAME statement specifies that the imputed values of MOMHOSP will be called MHOSP (with the _1, _2, _3, and _4 suffixes added for the multiple imputations.

The PRINT statement requests the DONORSTAT and MEANS print groups.

The OUTPUT statement includes the option IMPUTE=DEFAULT. This, used in conjunction with the ALL option on the IDVAR statement, requests that all variables from the input dataset be included on the output dataset MULTIMP, plus the following variables created in IMPUTE:

Variable	Definition
ORIGVAR1	Original imputation variable (values from MOMHOSP)
MHOSP_1	Imputed variable from Imputation #1
MHOSP_2	Imputed variable from Imputation #2
MHOSP_3	Imputed variable from Imputation #3
MHOSP_4	Imputed variable from Imputation #4
DONORID1	Donor ID from Imputation #1
DONORID2	Donor ID from Imputation #2
DONORID3	Donor ID from Imputation #3
DONORID4	Donor ID from Imputation #4

Exhibit 8. First Page of IMPUTE Results

```
                              S U D A A N
          Software for the Statistical Analysis of Correlated Data
         Copyright      Research Triangle Institute         June 2012
                              Release 11.0.0

The Impute Process has completed successfully.
Method: WSHD

Total Records Read from File: 953
Random Number Seed: 6537181

Total Imputation Classes: 6

Total Respondent Records: 932

Total Nonrespondent Records: 21
   Total Donor Records Imputation #1: 21
   Total Donor Records Imputation #2: 21
   Total Donor Records Imputation #3: 21
   Total Donor Records Imputation #4: 21
   Total Records Imputed: 21
```

The default summary output displayed above provides useful information about the IMPUTE run. The method of imputation is WSHD, the number of valid records read from the file is 953, the number of item respondent records is 932, the number of item nonrespondent records is 21, the number of donor records is 21, and the number of records imputed is 21 for each round of imputations. There are 6 imputation classes, since mother's race has 2 levels and smoking behavior has 3 levels (levels 1, 2, and missing). The number of respondent records plus the number of nonrespondent records equals the total records read from the file (932+21=953).

Exhibit 9. IMPUTE Results: DONORSTAT PRINT Group (Imputation #1)

```
Hot-Deck Multiple Imputations for MOMHOSP

Imputation #1
  Method = WSHD

by: Mother Race, Smoking behavior.
--------------------------------------------------
Mother Race
   Smoking
     behavi-   Item    Item               Missing
     or        Resp    Non-Resp   Donor   Data -
               Count   Count      Count   Post-imp
--------------------------------------------------
1
   Missing         6          0       0          0
   1             186          4       4          0
   2             272         12      12          0
2
   Missing        15          0       0          0
   1              39          1       1          0
   2             414          4       4          0
--------------------------------------------------
```

This table is the DONORSTAT print table for imputation #1. The rows in this table represent the complete cross-classification of the IMBY variables. In this example, there are 6 imputation classes (missing values for Smoking Behavior forms its own level). For each row, the table tells us the number of records considered to be respondents and nonrespondents, the number of donors contributing to imputation, and the number of records with missing data post-imputation. So for MOMRACE=1 and MOMSMK=1, there are 186 respondent records, 4 nonrespondent records (missing MOMHOSP), 4 donors contributing data, and no records with missing MOMHOSP after imputation.

The next 3 tables correspond to imputation rounds #2, 3, and 4.

Exhibit 10. IMPUTE Results: DONORSTAT PRINT Group (Imputation #2)

```
Hot-Deck Multiple Imputations for MOMHOSP

Imputation #2
  Method = WSHD

by: Mother Race, Smoking behavior.
--------------------------------------------------
Mother Race
   Smoking
     behavi-   Item    Item               Missing
     or        Resp    Non-Resp   Donor   Data -
               Count   Count      Count   Post-imp
--------------------------------------------------
1
   Missing         6          0       0          0
   1             186          4       4          0
   2             272         12      12          0
2
   Missing        15          0       0          0
   1              39          1       1          0
   2             414          4       4          0
--------------------------------------------------
```

Exhibit 11. IMPUTE Results: DONORSTAT PRINT Group (Imputation #3)

```
Hot-Deck Multiple Imputations for MOMHOSP

Imputation #3
  Method = WSHD

by: Mother Race, Smoking behavior.
-------------------------------------------------
Mother Race
   Smoking
     behavi-   Item    Item               Missing
     or        Resp    Non-Resp   Donor   Data -
               Count   Count      Count   Post-imp
-------------------------------------------------
1
   Missing         6          0       0          0
   1             186          4       4          0
   2             272         12      12          0
2
   Missing        15          0       0          0
   1              39          1       1          0
   2             414          4       4          0
-------------------------------------------------
```

Exhibit 12. IMPUTE Results: DONORSTAT PRINT Group (Imputation #4)

```
Hot-Deck Multiple Imputations for MOMHOSP

Imputation #4
  Method = WSHD

by: Mother Race, Smoking behavior.
-------------------------------------------------
Mother Race
   Smoking
     behavi-   Item    Item               Missing
     or        Resp    Non-Resp   Donor   Data -
               Count   Count      Count   Post-imp
-------------------------------------------------
1
   Missing         6          0       0          0
   1             186          4       4          0
   2             272         12      12          0
2
   Missing        15          0       0          0
   1              39          1       1          0
   2             414          4       4          0
-------------------------------------------------
```

Exhibit 13. IMPUTE Results: MEANS PRINT Group (Imputation #1)

```
Hot-Deck Multiple Imputations for MOMHOSP

Imputation #1
  Method = WSHD
  Length of hospital stay -- mother

by: Mother Race, Smoking behavior.
-------------------------------------------------------
Mother Race
   Smoking                                        Rel.
     behavi-   Mean -     Mean -     Abs.       Diff.
     or        Pre-imp    Post-imp   Diff.      (%)
-------------------------------------------------------
1
   Missing         1.77       1.77    0.0000    0.0000
   1               2.13       2.13    0.0013   -0.0622
   2               2.02       1.99    0.0322   -1.5943
2
   Missing         1.47       1.47    0.0000    0.0000
   1               2.27       2.26    0.0164   -0.7204
   2               2.35       2.35    0.0010    0.0429
-------------------------------------------------------
```

This table corresponds to Imputation Round #1. It tells us the weighted mean of MOMHOSP before vs. after hot-deck imputation, within each imputation class. If no imputation took place within a cell, the weighted mean is the same before vs. after imputation. The first and fourth rows show no change since there were no nonrespondents in these cells. The remainder of the cells only shows small changes after imputation.

Exhibit 14. IMPUTE Results: MEANS PRINT Group (Imputation #2)

```
Hot-Deck Multiple Imputations for MOMHOSP

Imputation #2
  Method = WSHD
  Length of hospital stay -- mother

by: Mother Race, Smoking behavior.
-------------------------------------------------------
Mother Race
   Smoking                                        Rel.
     behavi-   Mean -     Mean -     Abs.       Diff.
     or        Pre-imp    Post-imp   Diff.      (%)
-------------------------------------------------------
1
   Missing         1.77       1.77    0.0000    0.0000
   1               2.13       2.12    0.0057   -0.2688
   2               2.02       2.01    0.0101   -0.5020
2
   Missing         1.47       1.47    0.0000    0.0000
   1               2.27       2.32    0.0438    1.9287
   2               2.35       2.33    0.0121   -0.5147
-------------------------------------------------------
```

Exhibit 15. IMPUTE Results: MEANS PRINT Group (Imputation #3)

```
Hot-Deck Multiple Imputations for MOMHOSP

Imputation #3
  Method = WSHD
  Length of hospital stay -- mother

by: Mother Race, Smoking behavior.
-----------------------------------------------------
Mother Race
   Smoking                                    Rel.
     behavi-   Mean -     Mean -     Abs.     Diff.
     or        Pre-imp    Post-imp   Diff.    (%)
-----------------------------------------------------
1
   Missing         1.77       1.77    0.0000    0.0000
   1               2.13       2.13    0.0049    0.2315
   2               2.02       2.01    0.0147   -0.7268
2
   Missing         1.47       1.47    0.0000    0.0000
   1               2.27       2.44    0.1642    7.2269
   2               2.35       2.36    0.0131    0.5580
-----------------------------------------------------
```

Exhibit 16. IMPUTE Results: MEANS PRINT Group (Imputation #4)

```
Hot-Deck Multiple Imputations for MOMHOSP

Imputation #4
  Method = WSHD
  Length of hospital stay -- mother

by: Mother Race, Smoking behavior.
-----------------------------------------------------
Mother Race
   Smoking                                    Rel.
     behavi-   Mean -     Mean -     Abs.     Diff.
     or        Pre-imp    Post-imp   Diff.    (%)
-----------------------------------------------------
1
   Missing         1.77       1.77    0.0000    0.0000
   1               2.13       2.13    0.0067    0.3136
   2               2.02       2.01    0.0071   -0.3491
2
   Missing         1.47       1.47    0.0000    0.0000
   1               2.27       2.26    0.0164   -0.7204
   2               2.35       2.34    0.0003   -0.0112
-----------------------------------------------------
```

Exhibit 17. RECORDS Results: Contents of WSHD Imputed Dataset

```
                                S U D A A N
          Software for the Statistical Analysis of Correlated Data
         Copyright      Research Triangle Institute         June 2012
                                Release 11.0.0

SAS Record File OUT.MULTIMP
Variables
Name          Type       Output     SAS          Description
                         Format     Format
-------------------------------------------------------------
PROCNUM       Numeric    F2.                     Procedure Number
DONORID1      Numeric    F8.                     dnr1
DONORID2      Numeric    F8.                     dnr2
DONORID3      Numeric    F8.                     dnr3
DONORID4      Numeric    F8.                     dnr4
MHOSP_1       Numeric                            hsp1
MHOSP_2       Numeric                            hsp2
MHOSP_3       Numeric                            hsp3
MHOSP_4       Numeric                            hsp4
ID            Numeric    F5.                     ID
MARITAL2      Numeric                            Marital Status
BABYDOB       Numeric               MMDDYY8.     Birth date--Baby
BABYSEX       Numeric    F12.                    Sex of baby
BABHOSP       Numeric                            Length of hospital stay -- baby
MOMHOSP       Numeric                            momhos
BORN_US       Numeric                            Place of birth -- US
COUNTRY       Alpha      A20                     Place of birth -- outside of US
YRSINUS       Numeric                            Years in US
HHSIZE        Numeric                            HH size
MOMONWIC      Numeric                            Mom on WIC while pregnant
DOBMOM        Numeric               MMDDYY8.     Date of birth--Mom
MOINCOM       Numeric    F12.                    Family income
MOMSMK        Numeric                            momsmk
DURATION      Numeric    F3.                     Breastfeeding duration -- computed
MOLASTBF      Numeric                            Last breastfeeding month
MOSTOPBF      Numeric                            Month reported breastfeeding stopped
BRFDINIT      Numeric    F12.                    Breastfeeding Initiated
FRSTDAYS      Numeric    F3.                     Age of baby when formula started
SITE          Numeric    F12.                    Site
ANALWGT1      Numeric    F12.                    Analysis
STRATUM       Numeric    F12.                    STRATUM
PSU           Numeric    F12.                    PSU
BABYWGT       Numeric    F12.                    Baby Weight (ozs.)
EDUC          Numeric    F12.                    Education
BFEED         Numeric    F12.                    Breastfeeding Initiation
MOMRACE       Numeric    F12.                    momrac
AGE           Numeric                            AGE
MRTLSTAT      Numeric    F12.                    Marital Status
BRFDEND       Numeric    F12.                    Breastfeeding stopped (0-1-.)
ORIGVAL       Numeric                            Original Variable
```

This table shows the contents of the WSHD imputed dataset.

Exhibit 18. RECORDS Results: First Page of WSHD Imputed Dataset

WSHD Multiply Imputed Dataset

by: Observation Number.

Observation Number	ID	mom rac	mom smk	dnr1	dnr2	dnr3	dnr4	mom hos	hsp 1	hsp 2	hsp 3	hsp 4
1	4001	1	.	.	.	.	.	2	2	2	2	2
2	19136	1	.	.	.	.	.	2	2	2	2	2
3	21017	1	.	.	.	.	.	1	1	1	1	1
4	32019	1	.	.	.	.	.	1	1	1	1	1
5	40016	1	.	.	.	.	.	1	1	1	1	1
6	42125	1	.	.	.	.	.	4	4	4	4	4
7	1001	1	1	.	.	.	.	2	2	2	2	2
8	1009	1	1	.	.	.	.	2	2	2	2	2
9	1018	1	1	.	.	.	.	2	2	2	2	2
10	1020	1	1	.	.	.	.	3	3	3	3	3
11	1021	1	1	.	.	.	.	3	3	3	3	3
12	1022	1	1	.	.	.	.	3	3	3	3	3
13	1023	1	1	.	.	.	.	13	13	13	13	13
14	3012	1	1	.	.	.	.	0	0	0	0	0
15	3128	1	1	.	.	.	.	1	1	1	1	1
16	4003	1	1	.	.	.	.	0	0	0	0	0
17	4008	1	1	.	.	.	.	1	1	1	1	1
18	4022	1	1	.	.	.	.	1	1	1	1	1
19	4026	1	1	.	.	.	.	4	4	4	4	4
20	4127	1	1	.	.	.	.	3	3	3	3	3
21	4130	1	1	.	.	.	.	1	1	1	1	1
22	4133	1	1	.	.	.	.	1	1	1	1	1
23	4135	1	1	.	.	.	.	5	5	5	5	5
24	6008	1	1	.	.	.	.	2	2	2	2	2
25	6009	1	1	.	.	.	.	2	2	2	2	2
26	7122	1	1	.	.	.	.	2	2	2	2	2
27	8131	1	1	.	.	.	.	1	1	1	1	1
28	10129	1	1	.	.	.	.	1	1	1	1	1
29	**11013**	**1**	**1**	**12010**	**12014**	**8131**	**16016**	**.**	**3**	**1**	**1**	**2**
30	11014	1	1	.	.	.	.	1	1	1	1	1
31	11123	1	1	.	.	.	.	2	2	2	2	2
32	12004	1	1	.	.	.	.	1	1	1	1	1
33	12005	1	1	.	.	.	.	2	2	2	2	2
34	12009	1	1	.	.	.	.	1	1	1	1	1
35	12010	1	1	.	.	.	.	3	3	3	3	3
36	12013	1	1	.	.	.	.	1	1	1	1	1
37	12014	1	1	.	.	.	.	1	1	1	1	1
38	12117	1	1	.	.	.	.	2	2	2	2	2
39	13005	1	1	.	.	.	.	2	2	2	2	2
40	13126	1	1	.	.	.	.	2	2	2	2	2
41	14005	1	1	.	.	.	.	1	1	1	1	1
42	14015	1	1	.	.	.	.	1	1	1	1	1

There are 953 records on the output dataset MULTIMP. This is the first page of the output dataset. It is sorted by MOMRACE and MOMSMK. DONORID is missing except when imputation takes place. On those records, DONORID is the ID of the donor record. So for ID=11013 (record 29), MOMHOSP is missing. The random donors (one for each round of imputation) from the same race and smoking behavior levels are DONORID=12010 (who contributed a value of MOMHOSP=3), 12014 (contributed MOMHOSP=1), 8131 (contributed MOMHOSP=1), and 16016 (contributed MOMHOSP=2).

Exhibit 19. SAS FREQ Results: See How MOMHOSP Is Imputed (Imputation #1)

MOMHOSP	mhosp_1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	1	7	0.73	7	0.73
.	2	11	1.15	18	1.89
.	3	2	0.21	20	2.10
.	4	1	0.10	21	2.20
0	0	25	2.62	46	4.83
1	1	320	33.58	366	38.41
2	2	350	36.73	716	75.13
3	3	142	14.90	858	90.03
4	4	48	5.04	906	95.07
5	5	22	2.31	928	97.38
6	6	7	0.73	935	98.11
7	7	4	0.42	939	98.53
8	8	2	0.21	941	98.74
10	10	1	0.10	942	98.85
11	11	1	0.10	943	98.95
12	12	3	0.31	946	99.27
13	13	7	0.73	953	100.00

In Imputation Round #1, the 21 missing values for MOMHOSP were hot-deck imputed to either 1, 2, 3, or 4. All other values remained as is.

Exhibit 20. SAS FREQ Results: See How MOMHOSP Is Imputed (Imputation #2)

MOMHOSP	mhosp_2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	1	8	0.84	8	0.84
.	2	8	0.84	16	1.68
.	3	3	0.31	19	1.99
.	5	1	0.10	20	2.10
.	7	1	0.10	21	2.20
0	0	25	2.62	46	4.83
1	1	320	33.58	366	38.41
2	2	350	36.73	716	75.13
3	3	142	14.90	858	90.03
4	4	48	5.04	906	95.07
5	5	22	2.31	928	97.38
6	6	7	0.73	935	98.11
7	7	4	0.42	939	98.53
8	8	2	0.21	941	98.74
10	10	1	0.10	942	98.85
11	11	1	0.10	943	98.95
12	12	3	0.31	946	99.27
13	13	7	0.73	953	100.00

In Imputation Round #2, the 21 missing values for MOMHOSP were hot-deck imputed to either 1, 2, 3, 5, or 7. All other values remained as is.

Exhibit 21. SAS FREQ Results: See How MOMHOSP Is Imputed (Imputation #3)

MOMHOSP	mhosp_3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	1	5	0.52	5	0.52
.	2	8	0.84	13	1.36
.	3	4	0.42	17	1.78
.	4	3	0.31	20	2.10
.	5	1	0.10	21	2.20
0	0	25	2.62	46	4.83
1	1	320	33.58	366	38.41
2	2	350	36.73	716	75.13
3	3	142	14.90	858	90.03
4	4	48	5.04	906	95.07
5	5	22	2.31	928	97.38
6	6	7	0.73	935	98.11
7	7	4	0.42	939	98.53
8	8	2	0.21	941	98.74
10	10	1	0.10	942	98.85
11	11	1	0.10	943	98.95
12	12	3	0.31	946	99.27
13	13	7	0.73	953	100.00

In Imputation Round #3, the 21 missing values for MOMHOSP were hot-deck imputed to either 1, 2, 3, 4, or 5. All other values remained as is.

Exhibit 22. SAS FREQ Results: See How MOMHOSP Is Imputed (Imputation #4)

MOMHOSP	mhosp_4	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	0	1	0.10	1	0.10
.	1	5	0.52	6	0.63
.	2	10	1.05	16	1.68
.	3	2	0.21	18	1.89
.	4	2	0.21	20	2.10
.	7	1	0.10	21	2.20
0	0	25	2.62	46	4.83
1	1	320	33.58	366	38.41
2	2	350	36.73	716	75.13
3	3	142	14.90	858	90.03
4	4	48	5.04	906	95.07
5	5	22	2.31	928	97.38
6	6	7	0.73	935	98.11
7	7	4	0.42	939	98.53
8	8	2	0.21	941	98.74
10	10	1	0.10	942	98.85
11	11	1	0.10	943	98.95
12	12	3	0.31	946	99.27
13	13	7	0.73	953	100.00

In Imputation Round #4, the 21 missing values for MOMHOSP were hot-deck imputed to either 0, 1, 2, 3, 4, or 7. All other values remained as is.

One of the advantages of using WSHD imputation is that, in expectation, the distribution of item respondents is equivalent that generated from the imputation-completed data. The distribution of the imputed variables is printed using the FREQ procedure:

Exhibit 23. SAS FREQ Results: Distribution of MHOSP_1 (Imputation #1)

Imputation #1 for Imputed variable #1

mhosp_1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	25	2.62	25	2.62
1	327	34.31	352	36.94
2	361	37.88	713	74.82
3	144	15.11	857	89.93
4	49	5.14	906	95.07
5	22	2.31	928	97.38
6	7	0.73	935	98.11
7	4	0.42	939	98.53
8	2	0.21	941	98.74
10	1	0.10	942	98.85
11	1	0.10	943	98.95
12	3	0.31	946	99.27
13	7	0.73	953	100.00

Exhibit 24. SAS FREQ Results: Distribution of MHOSP_2 (Imputation #2)

Imputation #2 for Imputed variable #1

mhosp_2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	25	2.62	25	2.62
1	328	34.42	353	37.04
2	358	37.57	711	74.61
3	145	15.22	856	89.82
4	48	5.04	904	94.86
5	23	2.41	927	97.27
6	7	0.73	934	98.01
7	5	0.52	939	98.53
8	2	0.21	941	98.74
10	1	0.10	942	98.85
11	1	0.10	943	98.95
12	3	0.31	946	99.27
13	7	0.73	953	100.00

Exhibit 25. SAS FREQ Results: Distribution of MHOSP_3 (Imputation #3)

Imputation #3 for Imputed variable #1

mhosp_3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	25	2.62	25	2.62
1	325	34.10	350	36.73
2	358	37.57	708	74.29
3	146	15.32	854	89.61
4	51	5.35	905	94.96
5	23	2.41	928	97.38
6	7	0.73	935	98.11
7	4	0.42	939	98.53
8	2	0.21	941	98.74
10	1	0.10	942	98.85
11	1	0.10	943	98.95
12	3	0.31	946	99.27
13	7	0.73	953	100.00

Exhibit 26. SAS FREQ Results: Distribution of MHOSP_4 (Imputation #4)

Imputation #4 for Imputed variable #1

mhosp_4	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	26	2.73	26	2.73
1	325	34.10	351	36.83
2	360	37.78	711	74.61
3	144	15.11	855	89.72
4	50	5.25	905	94.96
5	22	2.31	927	97.27
6	7	0.73	934	98.01
7	5	0.52	939	98.53
8	2	0.21	941	98.74
10	1	0.10	942	98.85
11	1	0.10	943	98.95
12	3	0.31	946	99.27
13	7	0.73	953	100.00

These results show that there are no missing values in the imputation variable for any of the 4 multiple imputations. The output also shows that the imputation-completed distributions for all imputations are similar to the item respondent distributions shown at the beginning of this example.