



This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: Deductive Imputation

Contents

| | |
|---|----|
| General section..... | 3 |
| 1. Summary | 3 |
| 2. General description of the method | 3 |
| 2.1 Simple imputation rules..... | 3 |
| 2.2 The use of equality restrictions..... | 3 |
| 2.3 The use of non-negativity constraints..... | 5 |
| 3. Preparatory phase | 6 |
| 4. Examples – not tool specific..... | 6 |
| 4.1 Example: deductive imputation with equality restrictions | 6 |
| 4.2 Example: deductive imputation with equality and non-negativity restrictions | 8 |
| 5. Examples – tool specific..... | 9 |
| 6. Glossary..... | 10 |
| 7. References | 10 |
| Specific section..... | 11 |
| Interconnections with other modules..... | 12 |
| Administrative section..... | 14 |

General section

1. Summary

In general, imputations are predictions for the missing values, based on an explicit or implicit model. In some cases, however, imputations can also be derived directly from the values that were observed in the same record, using derivation rules that do not contain any parameters to be estimated, such as is the case in models.

For instance: suppose that businesses are asked in a survey to report their total turnover (T), turnover from the main activity (T_1), and turnover from side-line activities (T_2). If the value of one of these variables is missing, and if it may be assumed that the two observed values are correct, then the missing value can be calculated using the rule: $T_1 + T_2 = T$.

The above imputation rule is an example of *deductive* or *logical imputation*. In this imputation method, one identifies cases where it is possible, based on logical or mathematical relationships between the variables, to unambiguously derive the value of one or more missing variables from the values that were observed, under the assumption that the observed values are correct. For the missing variables for which this is possible, the uniquely derived value is the deductive imputation. The assumption that all observed values are correct requires that all erroneous values in the original data have been removed in a previous process step.

2. General description of the method

2.1 Simple imputation rules

Many deductive imputations can be performed using simple rules in ‘if-then’ form, for example:

if (*total labour costs* = ‘missing’ **and** *employees on the payroll* = 0)
then *total labour costs* := 0.

These rules are compiled by subject-matter experts. They can be applied with many different types of software.

In the remainder of this section, we discuss two methods that generate deductive imputations automatically based on restrictions that must be satisfied by the data. These methods work only for numerical data. A similar method for categorical data is given by De Waal et al. (2011, Section 9.2.4), but we do not discuss this method here, because business surveys usually involve numerical data.

2.2 The use of equality restrictions

A particularly rich source for deductive imputations is formed by the extensive systems of equations that should hold for Structural Business Statistics. A typical survey may involve around 100 variables with 30 equality restrictions. Most of these equality restrictions have the general form

$$\text{Total} = \text{Subtotal}_1 + \text{Subtotal}_2 + \dots + \text{Subtotal}_s. \quad (1)$$

If, in such a case, one of the subtotals or the total is missing, it is immediately clear with which value the missing variable should be imputed: there is a single equation with a single unknown, so a unique solution exists.

More generally, we may encounter several variables with missing values that are involved in several inter-related equality restrictions. This means we have a system of equations with multiple unknowns, for which it is not immediately clear whether the values of some missing variables are uniquely determined by this system, and, if so, what these unique values would be. However, this problem may be solved using techniques from linear algebra. Below we describe a method that automatically generates the deductive imputations from a given system of equations. This description is based on Pannekoek (2006).

Suppose that a record consists of p variables and that q linear equality restrictions apply to these p variables. The restrictions may be represented in the form

$$\mathbf{R}\mathbf{y} = \mathbf{b}, \quad (2)$$

where \mathbf{y} is a vector of length p with the variables, \mathbf{b} is a vector of length q with constant terms that appear in the restrictions, and \mathbf{R} is a $q \times p$ matrix in which each row represents one restriction and each column represents one variable. For example, consider a business survey where the operating income block consists of the following five variables:

| | |
|------------------------------------|-------|
| Net turnover from main activity | y_1 |
| Net turnover from other activities | y_2 |
| Total net turnover | y_3 |
| Total other operating income | y_4 |
| Total operating income | y_5 |

Two restrictions apply to these variables: $y_1 + y_2 = y_3$ and $y_3 + y_4 = y_5$. These restrictions can be formulated as a system of equations in the form (2) with $\mathbf{b} = \mathbf{0}$ and

$$\mathbf{R} = \begin{bmatrix} 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 \end{bmatrix}.$$

If the vector with variables \mathbf{y} consists of p_o observed values and p_m missing values, then, after a permutation of elements, this vector can be partitioned as $\mathbf{y} = (\mathbf{y}'_o, \mathbf{y}'_m)'$, in which \mathbf{y}_o is a vector of length p_o with the observed values and \mathbf{y}_m is a vector of length p_m with the missing values. If we partition \mathbf{R} accordingly, we can write:

$$[\mathbf{R}_o \quad \mathbf{R}_m] \begin{bmatrix} \mathbf{y}_o \\ \mathbf{y}_m \end{bmatrix} = \mathbf{b},$$

so that, say,

$$\mathbf{R}_m \mathbf{y}_m = \mathbf{b} - \mathbf{R}_o \mathbf{y}_o = \mathbf{a}. \quad (3)$$

Note that \mathbf{a} can be computed using only the observed values in the record. Thus, expression (3) is a system of linear equations that involves only the missing variables \mathbf{y}_m . The intention of deductive imputation is to derive as many missing values as possible from this system.

For a system of linear equations, one usually distinguishes between three cases:

- I) There are no solutions (the system is inconsistent);
- II) There is exactly one solution;

III) There are an infinite number of solutions.

For system (3) – assuming that the original restrictions in (2) do not contradict each other –, Case I can only occur if there are errors in the observed values. We assume here that all errors have been detected previously and replaced by missing values. Moreover, if this has been done using error localisation methodology as described in the module ‘Automatic Editing’, then it is certain that the missing values can be imputed in such a way that the restrictions are satisfied. Thus, under these assumptions, Case I cannot occur.

Case II occurs if \mathbf{R}_m is a matrix with rank equal to the number of missing values p_m . In the special case that \mathbf{R}_m is square, the unique \mathbf{y}_m that satisfies the restrictions is given by

$$\tilde{\mathbf{y}}_m = \mathbf{R}_m^{-1} \mathbf{a},$$

where \mathbf{R}_m^{-1} denotes the inverse matrix of \mathbf{R}_m . If \mathbf{R}_m is not square, we can still obtain a unique solution in this form after a suitable transformation of \mathbf{R}_m and \mathbf{a} to remove any linear dependent rows. Thus in Case II, all missing variables can be imputed deductively, since all missing values are uniquely determined by the system of equations and the observed values. This is an ideal situation.

In general, however, we will encounter Case III: there are an infinite number of solutions for \mathbf{y}_m . In this last case, it is still possible that some elements of \mathbf{y}_m have the same values in all possible solutions. These elements can be deductively imputed.

The general solution for \mathbf{y}_m to system (3) is given by (see, e.g., Rao, 1973, or Harville, 1997):

$$\tilde{\mathbf{y}}_m = \mathbf{R}_m^- \mathbf{a} + (\mathbf{R}_m^- \mathbf{R}_m - \mathbf{I}) \mathbf{z} = \mathbf{d} + \mathbf{Cz}, \quad (4)$$

where \mathbf{R}_m^- is a so-called generalised inverse of \mathbf{R}_m (i.e., a $p_m \times q$ matrix such that $\mathbf{R}_m \mathbf{R}_m^- \mathbf{R}_m = \mathbf{R}_m$), \mathbf{I} is the $p_m \times p_m$ identity matrix, and \mathbf{z} is an arbitrary vector of length p_m . Because \mathbf{z} can be chosen arbitrarily, expression (4) generates an infinite number of solutions for \mathbf{y}_m , except in the event that \mathbf{C} is a matrix of zeros, which can only occur in the above-mentioned Case II. However, if the matrix $\mathbf{C} = \mathbf{R}_m^- \mathbf{R}_m - \mathbf{I}$ contains rows with only zeros, then the corresponding elements of $\tilde{\mathbf{y}}_m$ are the same for all possible solutions, i.e., for each arbitrary choice of \mathbf{z} . These elements can thus be deductively imputed with the corresponding values of $\mathbf{d} = \mathbf{R}_m^- \mathbf{a}$. A straightforward procedure for computing a generalised inverse of any matrix is given by Greville (1959).

This method is illustrated by means of an example in Section 4.1.

2.3 The use of non-negativity constraints

Another possibility to perform deductive imputation is to use the fact that many variables have to be non-negative. Suppose, for example, that for the variables in restriction (1), only the value of Total and the values of Subtotal_1 and Subtotal_2 are observed, and suppose that these observed values satisfy:

$$\text{Total} = \text{Subtotal}_1 + \text{Subtotal}_2.$$

Clearly, the sum of the missing variables (Subtotal_3, ..., Subtotal_s) must be zero in this case. If the missing variables are not allowed to be negative, then this means that they can all be deductively imputed with zero.

To find these types of solutions in general, we again consider the system of equations $\mathbf{R}_m \mathbf{y}_m = \mathbf{a}$ found in (3). Suppose that there is an element a_j of \mathbf{a} that is equal to zero. For the corresponding row of \mathbf{R}_m , denoted by $\mathbf{r}'_{m,j}$, it must then hold that $\mathbf{r}'_{m,j} \mathbf{y}_m = 0$. Now, if, for all elements of \mathbf{y}_m that have non-zero coefficients in $\mathbf{r}'_{m,j}$, it is true that

- i) these elements y_{mi} must all be non-negative,
- ii) the non-zero coefficients in $\mathbf{r}'_{m,j}$ are either all negative or all positive,

then it is deduced that these elements of \mathbf{y}_m are all equal to zero.

The deductive imputations derived in this way for the missing values \mathbf{y}_m are therefore given by:

$$\tilde{y}_{mi} = 0, \text{ if } a_j = 0 \text{ and conditions i and ii are satisfied.}$$

This method is illustrated by means of an example in Section 4.2.

3. Preparatory phase

4. Examples – not tool specific

4.1 Example: deductive imputation with equality restrictions

To illustrate the method described in Section 2.2, we consider a fictitious survey with eleven variables that should satisfy five equality restrictions:

$$\left\{ \begin{array}{l} y_1 + y_2 = y_3 \\ y_2 = y_4 \\ y_5 + y_6 + y_7 = y_8 \\ y_3 + y_8 = y_9 \\ y_9 - y_{10} = y_{11} \end{array} \right.$$

This system of equations can be written in the form (2) with $\mathbf{b} = \mathbf{0}$ and

$$\mathbf{R} = \begin{bmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix}.$$

Suppose that we want to use deductive imputation to treat as many missing values as possible in the following incomplete record (where ‘-’ indicates a missing value):

$$\begin{array}{cccccccccccc}
y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} & y_{11} \\
154 & - & 166 & - & - & - & - & 25 & - & 204 & -
\end{array}$$

Making the appropriate partitions of \mathbf{R} and \mathbf{y} into observed and missing components, we compute

$$\mathbf{a} = -\mathbf{R}_o \mathbf{y}_o = - \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 154 \\ 166 \\ 25 \\ 204 \end{bmatrix} = \begin{bmatrix} 12 \\ 0 \\ 25 \\ -191 \\ 204 \end{bmatrix}$$

and thus obtain the following system $\mathbf{R}_m \mathbf{y}_m = \mathbf{a}$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} y_2 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_9 \\ y_{11} \end{bmatrix} = \begin{bmatrix} 12 \\ 0 \\ 25 \\ -191 \\ 204 \end{bmatrix}.$$

The following matrix \mathbf{R}_m^- satisfies $\mathbf{R}_m \mathbf{R}_m^- \mathbf{R}_m = \mathbf{R}_m$ and hence is a generalised inverse of \mathbf{R}_m :

$$\mathbf{R}_m^- = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 \end{bmatrix}.$$

Using this matrix in expression (4), we finally obtain:

$$\tilde{\mathbf{y}}_m = \mathbf{d} + \mathbf{Cz} = \begin{bmatrix} 12 \\ 12 \\ 25 \\ 0 \\ 0 \\ 191 \\ -13 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \end{bmatrix} = \begin{bmatrix} 12 \\ 12 \\ 25 + z_4 + z_5 \\ -z_4 \\ -z_5 \\ 191 \\ -13 \end{bmatrix}.$$

By inspection, it is seen that the first, second, sixth, and seventh rows of \mathbf{C} contain only zeros. This shows that we may deductively impute y_2 , y_4 , y_9 and y_{11} with the corresponding elements of \mathbf{d} . In this manner, we obtain the following partially imputed record:

$$\begin{array}{cccccccccccc}
y_1 & \tilde{y}_2 & y_3 & \tilde{y}_4 & y_5 & y_6 & y_7 & y_8 & \tilde{y}_9 & y_{10} & \tilde{y}_{11} \\
154 & 12 & 166 & 12 & - & - & - & 25 & 191 & 204 & -13
\end{array}$$

The remaining missing values in this example could not be imputed deductively. Imputations for these values have to be estimated by a non-deductive method. It should be noted that the accuracy of these estimated imputations may benefit from the fact that we have used deductive imputation, because more non-missing auxiliary values are now available.

4.2 Example: deductive imputation with equality and non-negativity restrictions

To illustrate the method described in Section 2.3, we consider the same set of restrictions as in the previous example, but with a different incomplete record:

$$\begin{array}{cccccccccccc} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} & y_{11} \\ 154 & - & 166 & - & 25 & - & - & 25 & - & 204 & - \end{array}$$

The only difference between this record and the record from Section 4.1 is that the value of y_5 is now also observed. In addition, all variables except y_{11} are now assumed to be non-negative.

Again partitioning \mathbf{R} and \mathbf{y} into observed and missing components, we obtain this time

$$\mathbf{a} = -\mathbf{R}_o \mathbf{y}_o = - \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 154 \\ 166 \\ 25 \\ 25 \\ 204 \end{bmatrix} = \begin{bmatrix} 12 \\ 0 \\ 0 \\ -191 \\ 204 \end{bmatrix}$$

and hence the following system $\mathbf{R}_m \mathbf{y}_m = \mathbf{a}$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} y_2 \\ y_4 \\ y_6 \\ y_7 \\ y_9 \\ y_{11} \end{bmatrix} = \begin{bmatrix} 12 \\ 0 \\ 0 \\ -191 \\ 204 \end{bmatrix}.$$

We note that the third row of this system states the following equation: $y_6 + y_7 = 0$. This equation has all the properties that we mentioned in Section 2.3: the right-hand-side equals zero, all coefficients have the same sign, and all variables involved have to be non-negative. Thus, we may deductively impute the values $\tilde{y}_6 = \tilde{y}_7 = 0$. The second row of the above system also represents an equation with right-hand-side equal to zero: $y_2 - y_4 = 0$. However, this equation contains both a positive and a negative coefficient, so it cannot be used to impute zeros in a deductive manner.

Since there are now two additional variables with non-missing values, we may update the partitions of \mathbf{R} and \mathbf{y} into observed and missing components. Using the method from Section 2.2 in the same way as before, we finally obtain the following, completely imputed record:

$$\begin{array}{cccccccccccc} y_1 & \tilde{y}_2 & y_3 & \tilde{y}_4 & y_5 & \tilde{y}_6 & \tilde{y}_7 & y_8 & \tilde{y}_9 & y_{10} & \tilde{y}_{11} \\ 154 & 12 & 166 & 12 & 25 & 0 & 0 & 25 & 191 & 204 & -13 \end{array}$$

5. Examples – tool specific

The R package `deducorrect`, which can be downloaded for free at <http://cran.r-project.org>, contains an implementation of the deductive imputation methods from Sections 2.2 and 2.3. To illustrate the use of `deducorrect`, we work out the two examples from Section 4 in R code.

First, we load the package:

```
> library(deducorrect)
```

Next, we create an object of type “editmatrix” containing the system of restrictions:

```
> E <- editmatrix( c("y1 + y2 == y3",
+                  "y2 == y4",
+                  "y5 + y6 + y7 == y8",
+                  "y3 + y8 == y9",
+                  "y9 - y10 == y11",
+                  "y1 >= 0", "y2 >= 0", "y3 >= 0",
+                  "y4 >= 0", "y5 >= 0", "y6 >= 0",
+                  "y7 >= 0", "y8 >= 0", "y9 >= 0",
+                  "y10 >= 0") )
```

We also have to read in the two records that we want to treat as a data frame:

```
> y <- data.frame( y1 = c(154, 154),
+                 y2 = c(NA, NA),
+                 y3 = c(166, 166),
+                 y4 = c(NA, NA),
+                 y5 = c(NA, 25),
+                 y6 = c(NA, NA),
+                 y7 = c(NA, NA),
+                 y8 = c(25, 25),
+                 y9 = c(NA, NA),
+                 y10 = c(204, 204),
+                 y11 = c(NA, NA) )
```

This produces the following data frame with two rows:

```
> y
  y1 y2  y3 y4 y5 y6 y7 y8 y9 y10 y11
1 154 NA 166 NA NA NA NA 25 NA 204  NA
2 154 NA 166 NA 25 NA NA 25 NA 204  NA
```

Deductive imputation may now be applied to these records by calling the function ‘`deduImpute`’ provided by the package:

```
> d <- deduImpute(E, y)
```

This command creates a list (named ‘`d`’ here) which contains the results of deductive imputation. We first check the status of each record:

```
> d$status
```

```

      status imputations
1  partial          4
2 corrected         6

```

This shows that the first record was partially imputed (with four imputations), while the second record was completely imputed (with six imputations). The imputed data itself is also stored in the list:

```

> d$corrected
      y1 y2  y3 y4 y5 y6 y7 y8  y9 y10 y11
1 154 12 166 12 NA NA NA 25 191 204 -13
2 154 12 166 12 25  0  0 25 191 204 -13

```

We refer to Van der Loo and De Jonge (2011) for more details on the `deducorrect` package.

6. Glossary

For definitions of terms used in this module, please refer to the separate “Glossary” provided as part of the handbook.

7. References

- De Waal, T., Pannekoek, J., and Scholtus, S. (2011), *Handbook of Statistical Data Editing and Imputation*. John Wiley & Sons, New Jersey.
- Greville, T. N. E. (1959), The Pseudoinverse of a Rectangular or Singular Matrix and Its Application to the Solution of Systems of Linear Equations. *SIAM Review* **1**, 38–43.
- Harville, D. A. (1997), *Matrix Algebra from a Statistician’s Perspective*. Springer-Verlag, New York.
- Pannekoek, J. (2006), Regression Imputation with Linear Equality Constraints on the Variables. Working Paper, UN/ECE Work Session on Statistical Data Editing, Bonn.
- Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, second edition. John Wiley & Sons, New York.
- Van der Loo, M. and de Jonge, E. (2011), Deductive Imputation with the `deducorrect` Package. Discussion Paper 201126, Statistics Netherlands, The Hague.

Specific section

8. Purpose of the method

Imputing missing values in microdata on logical grounds

9. Recommended use of the method

1. Deductive imputation is most effective when it is applied at the very beginning of the imputation process, after the removal of erroneous values, but before other forms of imputation have been used. In this way, other imputation methods have more non-missing auxiliary variables available, e.g., to estimate model parameters.

10. Possible disadvantages of the method

1. The method should be used, in principle, only for imputing values that can be derived with certainty from the observed values. In all other cases, it is usually better to use non-deductive methods, such as model-based imputation (see “Imputation – Model-Based Imputation”) or donor imputation (see “Imputation – Donor Imputation”).

11. Variants of the method

1. Deductive imputation by means of if-then rules specified by subject-matter specialists.
2. Automatic deductive imputation based on equality and non-negativity restrictions.

12. Input data

1. A data set containing microdata with missing values.

13. Logical preconditions

1. Missing values
 1. Allowed; in fact, the object of this method is to impute some of them.
2. Erroneous values
 1. Not allowed. Erroneous values have to be removed from the data in a previous step. They may be replaced by missing values.
3. Other quality related preconditions
 1. n/a
4. Other types of preconditions
 1. n/a

14. Tuning parameters

1. If relevant, a collection of restrictions (linear equations and – optionally – non-negativity constraints) for the microdata.

15. Recommended use of the individual variants of the method

1. Deductive imputation by means of if-then rules requires that subject-matter specialists design a collection of if-then rules beforehand.
2. Automatic deductive imputation is only possible if the data are restricted by equations and (optionally) non-negativity constraints. If such restrictions exist, then this variant is highly recommended.
3. Automatic deductive imputation based on equality and non-negativity restrictions requires software that can handle matrix computations. Not all survey-processing systems contain this type of functionality.
4. The two variants may be used in combination. In that case, it is recommended to start with automatic deductive imputation based on restrictions.

16. Output data

1. A data set containing partially imputed microdata, which is an updated version of the first input data set.

17. Properties of the output data

1. In the output data, all missing values in the input data have been imputed that could be derived on logical grounds from the observed values in the input data.
2. Typically, the output data still contain some missing values that have to be imputed by other methods.

18. Unit of input data suitable for the method

Incremental processing by record

19. User interaction - not tool specific

1. User interaction is not needed during an execution of deductive imputation.

20. Logging indicators

1. A list of (the number of) imputations per record, for future analyses.

21. Quality indicators of the output data

1. The fraction of missing values that have been imputed by the method.

22. Actual use of the method

1. ?

Interconnections with other modules

23. Themes that refer explicitly to this module

1. Imputation – Main Module

2. Imputation – Model-Based Imputation
3. Imputation – Donor Imputation

24. Related methods described in other modules

1. n/a

25. Mathematical techniques used by the method described in this module

1. (Generalised) matrix inversion

26. GSBPM phases where the method described in this module is used

1. GSBPM Sub-process 5.4: Impute

27. Tools that implement the method described in this module

1. R package `deducorrect`

28. Process step performed by the method

Imputation, i.e., determining and filling in new values for occurrences of missing or discarded values in a data file

Administrative section

29. Module code

Imputation-M-Deductive Imputation

30. Version history

| Version | Date | Description of changes | Author | Institute |
|---------|------------|--|-----------------|-------------------|
| 0.1 | 23-12-2011 | first version | Sander Scholtus | CBS (Netherlands) |
| 0.2 | 29-03-2011 | improvements based on Norwegian review | Sander Scholtus | CBS (Netherlands) |
| 0.2.1 | 04-03-2013 | adjusted to new template; minor improvements | Sander Scholtus | CBS (Netherlands) |
| 0.2.2 | 21-10-2013 | preliminary release | | |
| 1.0 | 26-03-2014 | final version within the Memobust project | | |
| | | | | |
| | | | | |
| | | | | |

31. Template version and print date

| | |
|-----------------------|-------------------------|
| Template version used | 1.0 p 4 d.d. 22-11-2012 |
| Print date | 21-3-2014 18:15 |