

3 Primary Scientific Question

Is mean and hot deck imputation capable of improving sensitivity of testing data files by increasing their fit to some analytical model.

4 Data Example

Four methods of data imputation are available; (1) mean imputation, (2) hot deck imputation, (3) regression imputation, (4) multiple imputations. In the current chapter the methods (1) and (2) will be given. The methods (3) and (4) are described in Statistics applied to clinical studies 5th edition, Chap. 22, Springer Heidelberg Germany, 2012, from the same authors as the current work. A condition for any type of data imputation is that the missing data are not clustered but randomly distributed in the data file. A data example a 35 patient study crossover study of the effects of age and traditional laxative efficacy (numbers of stools per month) on the performance of a novel laxative is in the underneath table.

New lax	Bisacodyl	Age
24,00	8,00	25,00
30,00	13,00	30,00
25,00	15,00	25,00
35,00	10,00	31,00
39,00	9,00	
30,00	10,00	33,00
27,00	8,00	22,00
14,00	5,00	18,00
39,00	13,00	14,00
42,00		30,00
41,00	11,00	36,00
38,00	11,00	30,00
39,00	12,00	27,00
37,00	10,00	38,00
47,00	18,00	40,00
	13,00	31,00
36,00	12,00	25,00
12,00	4,00	24,00
26,00	10,00	27,00
20,00	8,00	20,00
43,00	16,00	35,00
31,00	15,00	29,00
40,00	14,00	32,00
31,00		30,00

(continued)

New lax	Bisacodyl	Age
36,00	12,00	40,00
21,00	6,00	31,00
44,00	19,00	41,00
11,00	5,00	26,00
27,00	8,00	24,00
24,00	9,00	30,00
40,00	15,00	
32,00	7,00	31,00
10,00	6,00	23,00
37,00	14,00	43,00
19,00	7,00	30,00

Five values of the above table are missing. The underneath table gives the results of mean imputation of these data. The missing values are imputed by the mean values of the different variables.

New lax	Bisacodyl	Age
24,00	8,00	25,00
30,00	13,00	30,00
25,00	15,00	25,00
35,00	10,00	31,00
39,00	9,00	<u>29,00</u>
30,00	10,00	33,00
27,00	8,00	22,00
14,00	5,00	18,00
39,00	13,00	14,00
42,00	<u>11,00</u>	30,00
41,00	11,00	36,00
38,00	11,00	30,00
39,00	12,00	27,00
37,00	10,00	38,00
47,00	18,00	40,00
<u>30,00</u>	13,00	31,00
36,00	12,00	25,00
12,00	4,00	24,00
26,00	10,00	27,00
20,00	8,00	20,00
43,00	16,00	35,00
31,00	15,00	29,00
40,00	14,00	32,00
31,00	<u>11,00</u>	30,00
36,00	12,00	40,00
21,00	6,00	31,00
44,00	19,00	41,00

(continued)

New lax	Bisacodyl	Age
11,00	5,00	26,00
27,00	8,00	24,00
24,00	9,00	30,00
40,00	15,00	29,00
32,00	7,00	31,00
10,00	6,00	23,00
37,00	14,00	43,00
19,00	7,00	30,00

The underneath table gives the results of the second method, hot deck imputation. The missing data are imputed by those of the closest neighbour observed: the missing value is imputed with the value of an individual whose non-missing data are closest to those of the patient with the missing value.

New lax	Bisacodyl	Age
24,00	8,00	25,00
30,00	13,00	30,00
25,00	15,00	25,00
35,00	10,00	31,00
39,00	9,00	30,00
30,00	10,00	33,00
27,00	8,00	22,00
14,00	5,00	18,00
39,00	13,00	14,00
42,00	14,00	30,00
41,00	11,00	36,00
38,00	11,00	30,00
39,00	12,00	27,00
37,00	10,00	38,00
47,00	18,00	40,00
30,00	13,00	31,00
36,00	12,00	25,00
12,00	4,00	24,00
26,00	10,00	27,00
20,00	8,00	20,00
43,00	16,00	35,00
31,00	15,00	29,00
40,00	14,00	32,00
31,00	15,00	30,00
36,00	12,00	40,00
21,00	6,00	31,00
44,00	19,00	41,00
11,00	5,00	26,00
27,00	8,00	24,00

(continued)

New lax	Bisacodyl	Age
24,00	9,00	30,00
40,00	15,00	32,00
32,00	7,00	31,00
10,00	6,00	23,00
37,00	14,00	43,00
19,00	7,00	30,00

5 Conclusion

Imputed data are of course not real data, but constructed values that should increase the sensitivity of testing. Regression imputation is more sensitive than mean and hot deck imputation, but it often overstates sensitivity. Probably, the best method for data imputation is multiple imputations (4), because this method works as a device for representing missing data uncertainty. However, a pocket calculator is unable to perform the analysis, and a statistical software package like SPSS statistical software is required.

6 Note

More background, theoretical and mathematical information of missing data and data imputations is given in Statistics applied to clinical studies 5th edition, Chap. 22, Springer Heidelberg Germany, 2012, from the same authors.