Análisis y Tratamiento de Datos con R: Departamento de Matemática

*

6 de noviembre de 2017

Imputación de datos

Basados en: https://www.r-bloggers.com/missing-value-treatment/

```
# initialize the data
data ("BostonHousing", package="mlbench")
original <- BostonHousing  # backup original data

# Introduce missing values
set.seed(100)
BostonHousing[sample(1:nrow(BostonHousing), 40), "rad"] <- NA
BostonHousing[sample(1:nrow(BostonHousing), 40), "ptratio"] <- NA
```

¿Que datos están perdidos?

```
# Pattern of missing values
library(mice)
```

```
## Loading required package: lattice
```

```
md.pattern(BostonHousing)
```

```
##     crim zn indus chas nox rm age dis tax b lstat medv rad ptratio
## 431    1  1     1    1   1  1   1   1   1 1     1    1   1       1  0
##  35    1  1     1    1   1  1   1   1   1 1     1    1   0       1  1
##  35    1  1     1    1   1  1   1   1   1 1     1    1   1       0  1
##   5    1  1     1    1   1  1   1   1   1 1     1    1   0       0  2
##        0  0     0    0   0  0   0   0   0 0     0    0  40      40 80
```

Borrando observaciones

- Si se tiene suficientes observaciones, no se debería perder mucho
- No se introduce sesgos debido a alguna pérdida sistemática de datos (representatividad)

```
# Example
lm(medv ~ ptratio + rad, data=BostonHousing, na.action=na.omit)
```

*,

```
##
## Call:
## lm(formula = medv ~ ptratio + rad, data = BostonHousing, na.action = na.omit)
##
## Coefficients:
## (Intercept)      ptratio          rad
##     57.2724      -1.7836      -0.2035
```

Borrando variables

Depende de la importancia del poder prtedictor que pueda tener esa variable.

Imputacióm usando la media, mediana o moda

```
library(Hmisc)
```

```
## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

```
impute(BostonHousing$ptratio, mean)  # replace with mean
```

```
##        1        2        3        4        5        6        7        8
##  15.3000  17.8000  17.8000  18.7000  18.7000  18.7000  15.2000  15.2000
##        9       10       11       12       13       14       15       16
##  15.2000  15.2000  15.2000  15.2000  15.2000  21.0000  21.0000  21.0000
##       17       18       19       20       21       22       23       24
##  21.0000  21.0000  21.0000  21.0000  21.0000  21.0000  21.0000  21.0000
##       25       26       27       28       29       30       31       32
##  21.0000  21.0000  21.0000  21.0000  21.0000  21.0000  21.0000  21.0000
##       33       34       35       36       37       38       39       40
##  21.0000  21.0000  21.0000  19.2000  19.2000  19.2000  19.2000  18.3000
##       41       42       43       44       45       46       47       48
##  18.3000  17.9000 18.4676*  17.9000  17.9000  17.9000  17.9000  17.9000
##       49       50       51       52       53       54       55       56
##  17.9000  17.9000  16.8000  16.8000  16.8000  16.8000  21.1000  17.9000
##       57       58       59       60       61       62       63       64
##  17.3000  15.1000  19.7000  19.7000 18.4676*  19.7000  19.7000  19.7000
##       65       66       67       68       69       70       71       72
##  18.6000  16.1000  16.1000  18.9000  18.9000  18.9000  19.2000  19.2000
##       73       74       75       76       77       78       79       80
##  19.2000  19.2000  18.7000  18.7000  18.7000  18.7000  18.7000  18.7000
##       81       82       83       84       85       86       87       88
##  19.0000  19.0000  19.0000  19.0000  18.5000  18.5000  18.5000  18.5000
##       89       90       91       92       93       94       95       96
```

```
##  17.8000  17.8000  17.8000  17.8000  18.2000  18.2000  18.2000  18.0000
##       97       98       99      100      101      102      103      104
##  18.0000  18.0000 18.4676*  18.0000  20.9000  20.9000 18.4676* 18.4676*
##      105      106      107      108      109      110      111      112
##  20.9000  20.9000  20.9000  20.9000  20.9000  20.9000  20.9000  17.8000
##      113      114      115      116      117      118      119      120
## 18.4676*  17.8000  17.8000  17.8000 18.4676* 18.4676*  17.8000  17.8000
##      121      122      123      124      125      126      127      128
##  19.1000  19.1000  19.1000  19.1000 18.4676*  19.1000  19.1000  21.2000
##      129      130      131      132      133      134      135      136
##  21.2000  21.2000  21.2000  21.2000  21.2000  21.2000  21.2000 18.4676*
##      137      138      139      140      141      142      143      144
##  21.2000  21.2000  21.2000  21.2000  21.2000  21.2000  14.7000  14.7000
##      145      146      147      148      149      150      151      152
##  14.7000  14.7000  14.7000  14.7000  14.7000  14.7000  14.7000  14.7000
##      153      154      155      156      157      158      159      160
## 18.4676*  14.7000  14.7000 18.4676*  14.7000  14.7000  14.7000  14.7000
##      161      162      163      164      165      166      167      168
##  14.7000  14.7000  14.7000 18.4676*  14.7000  14.7000  14.7000 18.4676*
##      169      170      171      172      173      174      175      176
##  14.7000  14.7000  14.7000  14.7000 18.4676*  16.6000  16.6000  16.6000
##      177      178      179      180      181      182      183      184
##  16.6000  16.6000  16.6000  17.8000  17.8000  17.8000  17.8000  17.8000
##      185      186      187      188      189      190      191      192
##  17.8000  17.8000  17.8000  15.2000  15.2000  15.2000  15.2000  15.2000
##      193      194      195      196      197      198      199      200
##  15.2000  15.6000  15.6000  14.4000 18.4676*  12.6000  12.6000  17.0000
##      201      202      203      204      205      206      207      208
##  17.0000  14.7000  14.7000  14.7000  14.7000  18.6000  18.6000  18.6000
##      209      210      211      212      213      214      215      216
##  18.6000  18.6000  18.6000  18.6000  18.6000 18.4676* 18.4676*  18.6000
##      217      218      219      220      221      222      223      224
##  16.4000  16.4000 18.4676*  16.4000  17.4000  17.4000  17.4000  17.4000
##      225      226      227      228      229      230      231      232
##  17.4000 18.4676*  17.4000  17.4000  17.4000  17.4000  17.4000  17.4000
##      233      234      235      236      237      238      239      240
##  17.4000  17.4000  17.4000  17.4000  17.4000  17.4000  16.6000  16.6000
##      241      242      243      244      245      246      247      248
##  16.6000  16.6000  16.6000  16.6000  19.1000  19.1000 18.4676*  19.1000
##      249      250      251      252      253      254      255      256
##  19.1000  19.1000  19.1000  19.1000  19.1000  19.1000  16.4000  16.4000
##      257      258      259      260      261      262      263      264
##  15.9000  13.0000  13.0000  13.0000  13.0000  13.0000  13.0000  13.0000
##      265      266      267      268      269      270      271      272
##  13.0000  13.0000  13.0000  13.0000  13.0000  18.6000  18.6000 18.4676*
##      273      274      275      276      277      278      279      280
##  18.6000  18.6000  17.6000  17.6000  17.6000  17.6000  17.6000  14.9000
##      281      282      283      284      285      286      287      288
##  14.9000  14.9000  14.9000  13.6000  15.3000  15.3000  18.2000  16.6000
##      289      290      291      292      293      294      295      296
##  16.6000  16.6000 18.4676* 18.4676*  19.2000  16.0000 18.4676*  16.0000
##      297      298      299      300      301      302      303      304
##  16.0000  16.0000  14.8000  14.8000  14.8000  16.1000 18.4676*  16.1000
##      305      306      307      308      309      310      311      312
```

```
##  18.4000  18.4000  18.4000  18.4000  18.4000  18.4000  18.4000  18.4000
##      313      314      315      316      317      318      319      320
## 18.4676* 18.4676*  18.4000  18.4000  18.4000  18.4000  18.4000  18.4000
##      321      322      323      324      325      326      327      328
##  19.6000  19.6000  19.6000  19.6000  19.6000  19.6000 18.4676*  19.6000
##      329      330      331      332      333      334      335      336
##  16.9000  16.9000  16.9000 18.4676*  16.9000  20.2000  20.2000  20.2000
##      337      338      339      340      341      342      343      344
##  20.2000  20.2000  20.2000  20.2000  20.2000  15.5000  15.9000  17.6000
##      345      346      347      348      349      350      351      352
##  17.6000  18.8000  18.8000  17.9000  17.0000  19.7000  19.7000  18.3000
##      353      354      355      356      357      358      359      360
##  18.3000  17.0000  22.0000  22.0000  20.2000  20.2000  20.2000  20.2000
##      361      362      363      364      365      366      367      368
##  20.2000  20.2000  20.2000 18.4676*  20.2000  20.2000  20.2000  20.2000
##      369      370      371      372      373      374      375      376
##  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      377      378      379      380      381      382      383      384
##  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      385      386      387      388      389      390      391      392
##  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000 18.4676* 18.4676*
##      393      394      395      396      397      398      399      400
##  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      401      402      403      404      405      406      407      408
##  20.2000  20.2000 18.4676*  20.2000  20.2000  20.2000  20.2000  20.2000
##      409      410      411      412      413      414      415      416
##  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      417      418      419      420      421      422      423      424
## 18.4676*  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      425      426      427      428      429      430      431      432
##  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      433      434      435      436      437      438      439      440
##  20.2000  20.2000  20.2000  20.2000 18.4676*  20.2000  20.2000  20.2000
##      441      442      443      444      445      446      447      448
##  20.2000 18.4676*  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      449      450      451      452      453      454      455      456
##  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      457      458      459      460      461      462      463      464
##  20.2000 18.4676*  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      465      466      467      468      469      470      471      472
## 18.4676*  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      473      474      475      476      477      478      479      480
##  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      481      482      483      484      485      486      487      488
##  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000  20.2000
##      489      490      491      492      493      494      495      496
##  20.1000  20.1000  20.1000  20.1000  20.1000  19.2000  19.2000  19.2000
##      497      498      499      500      501      502      503      504
##  19.2000  19.2000  19.2000 18.4676*  19.2000  21.0000  21.0000  21.0000
##      505      506
##  21.0000  21.0000
```

```
impute(BostonHousing$ptratio, median)  # median
```

```
##      1      2      3      4      5      6      7      8      9     10     11     12
##   15.3   17.8   17.8   18.7   18.7   18.7   15.2   15.2   15.2   15.2   15.2   15.2
##     13     14     15     16     17     18     19     20     21     22     23     24
##   15.2   21.0   21.0   21.0   21.0   21.0   21.0   21.0   21.0   21.0   21.0   21.0
##     25     26     27     28     29     30     31     32     33     34     35     36
##   21.0   21.0   21.0   21.0   21.0   21.0   21.0   21.0   21.0   21.0   21.0   19.2
##     37     38     39     40     41     42     43     44     45     46     47     48
##   19.2   19.2   19.2   18.3   18.3   17.9 19.1*   17.9   17.9   17.9   17.9   17.9
##     49     50     51     52     53     54     55     56     57     58     59     60
##   17.9   17.9   16.8   16.8   16.8   16.8   21.1   17.9   17.3   15.1   19.7   19.7
##     61     62     63     64     65     66     67     68     69     70     71     72
##  19.1*   19.7   19.7   19.7   18.6   16.1   16.1   18.9   18.9   18.9   19.2   19.2
##     73     74     75     76     77     78     79     80     81     82     83     84
##   19.2   19.2   18.7   18.7   18.7   18.7   18.7   18.7   19.0   19.0   19.0   19.0
##     85     86     87     88     89     90     91     92     93     94     95     96
##   18.5   18.5   18.5   18.5   17.8   17.8   17.8   17.8   18.2   18.2   18.2   18.0
##     97     98     99    100    101    102    103    104    105    106    107    108
##   18.0   18.0 19.1*   18.0   20.9   20.9 19.1* 19.1*   20.9   20.9   20.9   20.9
##    109    110    111    112    113    114    115    116    117    118    119    120
##   20.9   20.9   20.9   17.8 19.1*   17.8   17.8   17.8 19.1* 19.1*   17.8   17.8
##    121    122    123    124    125    126    127    128    129    130    131    132
##   19.1   19.1   19.1   19.1 19.1*   19.1   19.1   21.2   21.2   21.2   21.2   21.2
##    133    134    135    136    137    138    139    140    141    142    143    144
##   21.2   21.2   21.2 19.1*   21.2   21.2   21.2   21.2   21.2   21.2   14.7   14.7
##    145    146    147    148    149    150    151    152    153    154    155    156
##   14.7   14.7   14.7   14.7   14.7   14.7   14.7   14.7 19.1*   14.7   14.7 19.1*
##    157    158    159    160    161    162    163    164    165    166    167    168
##   14.7   14.7   14.7   14.7   14.7   14.7   14.7 19.1*   14.7   14.7   14.7 19.1*
##    169    170    171    172    173    174    175    176    177    178    179    180
##   14.7   14.7   14.7   14.7 19.1*   16.6   16.6   16.6   16.6   16.6   16.6   17.8
##    181    182    183    184    185    186    187    188    189    190    191    192
##   17.8   17.8   17.8   17.8   17.8   17.8   17.8   15.2   15.2   15.2   15.2   15.2
##    193    194    195    196    197    198    199    200    201    202    203    204
##   15.2   15.6   15.6   14.4 19.1*   12.6   12.6   17.0   17.0   14.7   14.7   14.7
##    205    206    207    208    209    210    211    212    213    214    215    216
##   14.7   18.6   18.6   18.6   18.6   18.6   18.6   18.6   18.6 19.1* 19.1*   18.6
##    217    218    219    220    221    222    223    224    225    226    227    228
##   16.4   16.4 19.1*   16.4   17.4   17.4   17.4   17.4   17.4 19.1*   17.4   17.4
##    229    230    231    232    233    234    235    236    237    238    239    240
##   17.4   17.4   17.4   17.4   17.4   17.4   17.4   17.4   17.4   17.4   16.6   16.6
##    241    242    243    244    245    246    247    248    249    250    251    252
##   16.6   16.6   16.6   16.6   19.1   19.1 19.1*   19.1   19.1   19.1   19.1   19.1
##    253    254    255    256    257    258    259    260    261    262    263    264
##   19.1   19.1   16.4   16.4   15.9   13.0   13.0   13.0   13.0   13.0   13.0   13.0
##    265    266    267    268    269    270    271    272    273    274    275    276
##   13.0   13.0   13.0   13.0   13.0   18.6   18.6 19.1*   18.6   18.6   17.6   17.6
##    277    278    279    280    281    282    283    284    285    286    287    288
##   17.6   17.6   17.6   14.9   14.9   14.9   14.9   13.6   15.3   15.3   18.2   16.6
##    289    290    291    292    293    294    295    296    297    298    299    300
##   16.6   16.6 19.1* 19.1*   19.2   16.0 19.1*   16.0   16.0   16.0   14.8   14.8
##    301    302    303    304    305    306    307    308    309    310    311    312
##   14.8   16.1 19.1*   16.1   18.4   18.4   18.4   18.4   18.4   18.4   18.4   18.4
##    313    314    315    316    317    318    319    320    321    322    323    324
##  19.1* 19.1*   18.4   18.4   18.4   18.4   18.4   18.4   19.6   19.6   19.6   19.6
```

```
##   325   326   327   328   329   330   331   332   333   334   335   336
##  19.6  19.6 19.1*  19.6  16.9  16.9  16.9 19.1*  16.9  20.2  20.2  20.2
##   337   338   339   340   341   342   343   344   345   346   347   348
##  20.2  20.2  20.2  20.2  20.2  15.5  15.9  17.6  17.6  18.8  18.8  17.9
##   349   350   351   352   353   354   355   356   357   358   359   360
##  17.0  19.7  19.7  18.3  18.3  17.0  22.0  22.0  20.2  20.2  20.2  20.2
##   361   362   363   364   365   366   367   368   369   370   371   372
##  20.2  20.2  20.2 19.1*  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2
##   373   374   375   376   377   378   379   380   381   382   383   384
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2
##   385   386   387   388   389   390   391   392   393   394   395   396
##  20.2  20.2  20.2  20.2  20.2  20.2 19.1* 19.1*  20.2  20.2  20.2  20.2
##   397   398   399   400   401   402   403   404   405   406   407   408
##  20.2  20.2  20.2  20.2  20.2  20.2 19.1*  20.2  20.2  20.2  20.2  20.2
##   409   410   411   412   413   414   415   416   417   418   419   420
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2 19.1*  20.2  20.2  20.2
##   421   422   423   424   425   426   427   428   429   430   431   432
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2
##   433   434   435   436   437   438   439   440   441   442   443   444
##  20.2  20.2  20.2  20.2 19.1*  20.2  20.2  20.2  20.2 19.1*  20.2  20.2
##   445   446   447   448   449   450   451   452   453   454   455   456
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2
##   457   458   459   460   461   462   463   464   465   466   467   468
##  20.2 19.1*  20.2  20.2  20.2  20.2  20.2  20.2 19.1*  20.2  20.2  20.2
##   469   470   471   472   473   474   475   476   477   478   479   480
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2
##   481   482   483   484   485   486   487   488   489   490   491   492
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.1  20.1  20.1  20.1
##   493   494   495   496   497   498   499   500   501   502   503   504
##  20.1  19.2  19.2  19.2  19.2  19.2  19.2 19.1*  19.2  21.0  21.0  21.0
##   505   506
##  21.0  21.0
```

```
impute(BostonHousing$ptratio, 20)  # replace specific number
```

```
##     1     2     3     4     5     6     7     8     9    10    11    12
##  15.3  17.8  17.8  18.7  18.7  18.7  15.2  15.2  15.2  15.2  15.2  15.2
##    13    14    15    16    17    18    19    20    21    22    23    24
##  15.2  21.0  21.0  21.0  21.0  21.0  21.0  21.0  21.0  21.0  21.0  21.0
##    25    26    27    28    29    30    31    32    33    34    35    36
##  21.0  21.0  21.0  21.0  21.0  21.0  21.0  21.0  21.0  21.0  21.0  19.2
##    37    38    39    40    41    42    43    44    45    46    47    48
##  19.2  19.2  19.2  18.3  18.3  17.9 20.0*  17.9  17.9  17.9  17.9  17.9
##    49    50    51    52    53    54    55    56    57    58    59    60
##  17.9  17.9  16.8  16.8  16.8  16.8  21.1  17.9  17.3  15.1  19.7  19.7
##    61    62    63    64    65    66    67    68    69    70    71    72
## 20.0*  19.7  19.7  19.7  18.6  16.1  16.1  18.9  18.9  18.9  19.2  19.2
##    73    74    75    76    77    78    79    80    81    82    83    84
##  19.2  19.2  18.7  18.7  18.7  18.7  18.7  18.7  19.0  19.0  19.0  19.0
##    85    86    87    88    89    90    91    92    93    94    95    96
##  18.5  18.5  18.5  18.5  17.8  17.8  17.8  17.8  18.2  18.2  18.2  18.0
##    97    98    99   100   101   102   103   104   105   106   107   108
##  18.0  18.0 20.0*  18.0  20.9  20.9 20.0* 20.0*  20.9  20.9  20.9  20.9
##   109   110   111   112   113   114   115   116   117   118   119   120
```

```
##  20.9  20.9  20.9  17.8 20.0*  17.8  17.8  17.8 20.0* 20.0*  17.8  17.8
##   121   122   123   124   125   126   127   128   129   130   131   132
##  19.1  19.1  19.1  19.1 20.0*  19.1  19.1  21.2  21.2  21.2  21.2  21.2
##   133   134   135   136   137   138   139   140   141   142   143   144
##  21.2  21.2  21.2 20.0*  21.2  21.2  21.2  21.2  21.2  21.2  14.7  14.7
##   145   146   147   148   149   150   151   152   153   154   155   156
##  14.7  14.7  14.7  14.7  14.7  14.7  14.7  14.7 20.0*  14.7  14.7 20.0*
##   157   158   159   160   161   162   163   164   165   166   167   168
##  14.7  14.7  14.7  14.7  14.7  14.7  14.7 20.0*  14.7  14.7  14.7 20.0*
##   169   170   171   172   173   174   175   176   177   178   179   180
##  14.7  14.7  14.7  14.7 20.0*  16.6  16.6  16.6  16.6  16.6  16.6  17.8
##   181   182   183   184   185   186   187   188   189   190   191   192
##  17.8  17.8  17.8  17.8  17.8  17.8  17.8  15.2  15.2  15.2  15.2  15.2
##   193   194   195   196   197   198   199   200   201   202   203   204
##  15.2  15.6  15.6  14.4 20.0*  12.6  12.6  17.0  17.0  14.7  14.7  14.7
##   205   206   207   208   209   210   211   212   213   214   215   216
##  14.7  18.6  18.6  18.6  18.6  18.6  18.6  18.6  18.6 20.0* 20.0*  18.6
##   217   218   219   220   221   222   223   224   225   226   227   228
##  16.4  16.4 20.0*  16.4  17.4  17.4  17.4  17.4  17.4 20.0*  17.4  17.4
##   229   230   231   232   233   234   235   236   237   238   239   240
##  17.4  17.4  17.4  17.4  17.4  17.4  17.4  17.4  17.4  17.4  16.6  16.6
##   241   242   243   244   245   246   247   248   249   250   251   252
##  16.6  16.6  16.6  16.6  19.1  19.1 20.0*  19.1  19.1  19.1  19.1  19.1
##   253   254   255   256   257   258   259   260   261   262   263   264
##  19.1  19.1  16.4  16.4  15.9  13.0  13.0  13.0  13.0  13.0  13.0  13.0
##   265   266   267   268   269   270   271   272   273   274   275   276
##  13.0  13.0  13.0  13.0  13.0  18.6  18.6 20.0*  18.6  18.6  17.6  17.6
##   277   278   279   280   281   282   283   284   285   286   287   288
##  17.6  17.6  17.6  14.9  14.9  14.9  14.9  13.6  15.3  15.3  18.2  16.6
##   289   290   291   292   293   294   295   296   297   298   299   300
##  16.6  16.6 20.0* 20.0*  19.2  16.0 20.0*  16.0  16.0  16.0  14.8  14.8
##   301   302   303   304   305   306   307   308   309   310   311   312
##  14.8  16.1 20.0*  16.1  18.4  18.4  18.4  18.4  18.4  18.4  18.4  18.4
##   313   314   315   316   317   318   319   320   321   322   323   324
## 20.0* 20.0*  18.4  18.4  18.4  18.4  18.4  18.4  19.6  19.6  19.6  19.6
##   325   326   327   328   329   330   331   332   333   334   335   336
##  19.6  19.6 20.0*  19.6  16.9  16.9  16.9 20.0*  16.9  20.2  20.2  20.2
##   337   338   339   340   341   342   343   344   345   346   347   348
##  20.2  20.2  20.2  20.2  20.2  15.5  15.9  17.6  17.6  18.8  18.8  17.9
##   349   350   351   352   353   354   355   356   357   358   359   360
##  17.0  19.7  19.7  18.3  18.3  17.0  22.0  22.0  20.2  20.2  20.2  20.2
##   361   362   363   364   365   366   367   368   369   370   371   372
##  20.2  20.2  20.2 20.0*  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2
##   373   374   375   376   377   378   379   380   381   382   383   384
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2
##   385   386   387   388   389   390   391   392   393   394   395   396
##  20.2  20.2  20.2  20.2  20.2  20.2 20.0* 20.0*  20.2  20.2  20.2  20.2
##   397   398   399   400   401   402   403   404   405   406   407   408
##  20.2  20.2  20.2  20.2  20.2  20.2 20.0*  20.2  20.2  20.2  20.2  20.2
##   409   410   411   412   413   414   415   416   417   418   419   420
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2 20.0*  20.2  20.2  20.2
##   421   422   423   424   425   426   427   428   429   430   431   432
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2
##   433   434   435   436   437   438   439   440   441   442   443   444
```

```
##  20.2  20.2  20.2  20.2 20.0*  20.2  20.2  20.2  20.2 20.0*  20.2  20.2
##   445   446   447   448   449   450   451   452   453   454   455   456
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2
##   457   458   459   460   461   462   463   464   465   466   467   468
##  20.2 20.0*  20.2  20.2  20.2  20.2  20.2  20.2 20.0*  20.2  20.2  20.2
##   469   470   471   472   473   474   475   476   477   478   479   480
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2
##   481   482   483   484   485   486   487   488   489   490   491   492
##  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.2  20.1  20.1  20.1  20.1
##   493   494   495   496   497   498   499   500   501   502   503   504
##  20.1  19.2  19.2  19.2  19.2  19.2  19.2 20.0*  19.2  21.0  21.0  21.0
##   505   506
##  21.0  21.0
```

```
# or if you want to impute manually
## BostonHousing$ptratio[is.na(BostonHousing$ptratio)] <- mean(BostonHousing$ptratio, na.rm = T)  # not
```

Calculemos la precisión cando se imputa con la media:

```
library(DMwR)
```

```
## Loading required package: grid
```

```
actuals <- original$ptratio[is.na(BostonHousing$ptratio)]
predicteds <- rep(mean(BostonHousing$ptratio, na.rm=T), length(actuals))
regr.eval(actuals, predicteds)
```

```
##        mae        mse       rmse       mape
## 1.62324034 4.19306071 2.04769644 0.09545664
```

Imputación basada en predicciones:

kNN Imputación

```
library(DMwR)
knnOutput <- knnImputation(BostonHousing[, !names(BostonHousing) %in% "medv"],k =5)  # perform knn impu
anyNA(knnOutput)
```

```
## [1] FALSE
```

```
#> FALSE
```

Calculemos la precisión

```
actuals <- original$ptratio[is.na(BostonHousing$ptratio)]
predicteds <- knnOutput[is.na(BostonHousing$ptratio), "ptratio"]
regr.eval(actuals, predicteds)
```

```
##        mae        mse       rmse       mape
## 0.75909106 1.32301168 1.15022245 0.04398371
```

rpart

Solo una de las variables redictoras tiene `NA` y puede tratar con variables factor

```
library(rpart)
class_mod <- rpart(rad ~ . - medv, data=BostonHousing[!is.na(BostonHousing$rad), ], method="class", na.a
anova_mod <- rpart(ptratio ~ . - medv, data=BostonHousing[!is.na(BostonHousing$ptratio), ], method="ano
rad_pred <- predict(class_mod, BostonHousing[is.na(BostonHousing$rad), ])
ptratio_pred <- predict(anova_mod, BostonHousing[is.na(BostonHousing$ptratio), ])
```

Calculemos la precisión para `ptratio`

```
actuals <- original$ptratio[is.na(BostonHousing$ptratio)]
predicteds <- ptratio_pred
regr.eval(actuals, predicteds)
```

```
##        mae        mse       rmse       mape
## 0.71061673 0.99693845 0.99846805 0.04099908
```

Para `rad`:

```
actuals <- original$rad[is.na(BostonHousing$rad)]
predicteds <- as.numeric(colnames(rad_pred)[apply(rad_pred, 1, which.max)])
mean(actuals != predicteds)
```

```
## [1] 0.25
```

MICE (Multivariate Imputation by Chained Equations)

Produce multiples copias completas de los datos, con valores diferentes sólo para los valores perdidos:

```
library(mice)
library(randomForest)
```

```
## randomForest 4.6-12

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:Hmisc':
##
##     combine

## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
miceMod <- mice(BostonHousing[, !names(BostonHousing) %in% "medv"], method="rf")  # perform mice imputa
```

```
##
##  iter imp variable
##   1   1  rad  ptratio
##   1   2  rad  ptratio
##   1   3  rad  ptratio
##   1   4  rad  ptratio
##   1   5  rad  ptratio
```

```
##   2   1  rad  ptratio
##   2   2  rad  ptratio
##   2   3  rad  ptratio
##   2   4  rad  ptratio
##   2   5  rad  ptratio
##   3   1  rad  ptratio
##   3   2  rad  ptratio
##   3   3  rad  ptratio
##   3   4  rad  ptratio
##   3   5  rad  ptratio
##   4   1  rad  ptratio
##   4   2  rad  ptratio
##   4   3  rad  ptratio
##   4   4  rad  ptratio
##   4   5  rad  ptratio
##   5   1  rad  ptratio
##   5   2  rad  ptratio
##   5   3  rad  ptratio
##   5   4  rad  ptratio
##   5   5  rad  ptratio
```

```
miceOutput <- complete(miceMod)  # generate the completed data.
anyNA(miceOutput)
```

```
## [1] FALSE
```

Para evaluar la precisión para `ptratio`:

```
actuals <- original$ptratio[is.na(BostonHousing$ptratio)]
predicteds <- miceOutput[is.na(BostonHousing$ptratio), "ptratio"]
regr.eval(actuals, predicteds)
```

```
##        mae        mse       rmse       mape
## 0.88250000 3.31225000 1.81995879 0.05378101
```

Y para `rad`:

```
actuals <- original$rad[is.na(BostonHousing$rad)]
predicteds <- miceOutput[is.na(BostonHousing$rad), "rad"]
mean(actuals != predicteds)
```

```
## [1] 0.225
```