Imputation-RegressionBased

SP

14/11/2019

```
ds<-read.csv("fitness.csv")
ds
```

```
##    Oxygen RunTime RunPulse
## 1  44.609   11.37      178
## 2  45.313   10.07      185
## 3  54.297    8.65      156
## 4  59.571      NA     <NA>
## 5  49.874    9.22     NA
## 6  44.811   11.63      176
## 7  45.681   11.95      176
## 8  49.091   10.85     <NA>
## 9  39.442   13.08      174
## 10 60.055    8.63      170
## 11 50.541      NA     NA
## 12 37.388   14.03      186
## 13 44.754   11.12      176
## 14 47.273      NA     <NA>
## 15 51.855   10.33      166
## 16 49.156    8.95      180
## 17 40.836   10.95      168
## 18 46.672   10.00     <NA>
## 19 46.774   10.25      NA
## 20 50.388   10.08      168
## 21 39.407   12.63      174
## 22 46.080   11.17      156
## 23 45.441    9.63      164
## 24 54.625    8.92      146
## 25 45.118   11.08     NA
## 26 39.203   12.88      168
## 27 45.790   10.47      186
## 28 50.545    9.93      148
## 29 48.673    9.40      186
## 30 47.920   11.50      170
## 31 47.467   10.50      170
```

RunPulse was read as the factor, so we will convert it into int

```
ds$RunPulse = as.double(ds$RunPulse)
ds
```

```
##    Oxygen RunTime RunPulse
## 1  44.609   11.37       10
## 2  45.313   10.07       12
## 3  54.297    8.65        3
## 4  59.571      NA       NA
```

```
## 5  49.874    9.22       15
## 6  44.811   11.63        9
## 7  45.681   11.95        9
## 8  49.091   10.85       NA
## 9  39.442   13.08        8
## 10 60.055    8.63        7
## 11 50.541      NA       15
## 12 37.388   14.03       13
## 13 44.754   11.12        9
## 14 47.273      NA       NA
## 15 51.855   10.33        5
## 16 49.156    8.95       11
## 17 40.836   10.95        6
## 18 46.672   10.00       NA
## 19 46.774   10.25       14
## 20 50.388   10.08        6
## 21 39.407   12.63        8
## 22 46.080   11.17        3
## 23 45.441    9.63        4
## 24 54.625    8.92        1
## 25 45.118   11.08       15
## 26 39.203   12.88        6
## 27 45.790   10.47       13
## 28 50.545    9.93        2
## 29 48.673    9.40       13
## 30 47.920   11.50        7
## 31 47.467   10.50        7
```

Missing values

Oxygen - has complete data

RunTime - has three observations missing

RunPulse - has three observations (4, 11, 14) missing together with RunTime and five on its own (5, 8, 18,19, 25)

Implementation of the regression as follows

RunTime - on Oxygen to predict missing observations 4, 11, 14

```
RunTime_on_Oxygen<-lm(RunTime~Oxygen,data=ds)
new_ds_1<-data.frame(Oxygen=c(59.571,50.541,47.273))
predict(RunTime_on_Oxygen,new_ds_1)
```

```
##         1         2         3
##  7.733491  9.827755 10.585679
```

RunPulse - on Oxygen to predict missing observations 4, 11, 14

```
RunPulse_on_Oxygen<-lm(RunPulse~Oxygen,data=ds)
predict(RunPulse_on_Oxygen,new_ds_1)
```

```
##        1        2        3
## 6.126922 7.856391 8.482294
```

RunPulse - on Oxygen and RunTime to predict missing observations 5, 8, 18, 19, 25

```
RunPulse_on_Oxygen_RunTime<-lm(RunPulse~RunTime+Oxygen,data=ds)
new_ds_1<-data.frame(Oxygen=c(49.874,49.091,46.672,46.774,45.118),RunTime=c(9.22,10.85,10.00,10.25,11.08
predict(RunPulse_on_Oxygen_RunTime,new_ds_1)
```

```
##        1        2        3        4        5
## 8.357936 7.060435 9.068702 8.765094 8.696585
```

```
new_ds_1
```

```
##   Oxygen RunTime
## 1 49.874    9.22
## 2 49.091   10.85
## 3 46.672   10.00
## 4 46.774   10.25
## 5 45.118   11.08
```