Relatório 1 - Regressão

Flavio Margarito Martins de Barros

14/05/2022

Conjunto de dados

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```r
## Carregando os pacotes
require(readxl)
require(corrplot)
require(psych)
require(kableExtra)
require(caret)
require(GGally)
require(Hmisc)
```

```r
## Lendo o banco de dados
dados <- read_excel(path = "Concrete_Data.xls", sheet = 1)

## Trocando os nomes das variáveis para o português
colnames(dados) <- c("cimento", "escoria", "cinza", "agua", "super_plastificante",
                     "agregador_grosso", "agregador_fino", "idade", "forca_compressiva")
```

```r
## Sumario dos dados
d <- Hmisc::describe(dados)
```

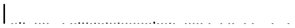
dados

9 Variables 1030 Observations

cimento

```
   n  missing  distinct   Info   Mean    Gmd    .05    .10    .25    .50    .75    .90    .95
1030        0       280      1  281.2  118.5  143.7  153.5  192.4  272.9  350.0  425.0  480.0

lowest : 102.0 108.3 116.0 122.6 132.0, highest: 522.0 525.0 528.0 531.3 540.0
```

escoria

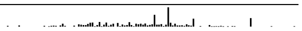


```
   n  missing  distinct   Info   Mean    Gmd  .05  .10  .25   .50   .75   .90   .95
1030        0       187  0.907   73.9  91.71  0.0  0.0  0.0  22.0 142.9 192.0 236.0

lowest :   0.00   0.02  11.00  13.61  15.00, highest: 290.20 305.30 316.10 342.10 359.40
```

cinza

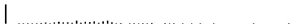
```
   n  missing  distinct   Info   Mean    Gmd  .05  .10  .25  .50   .75   .90   .95
1030        0       163  0.834  54.19  67.08  0.0  0.0  0.0  0.0 118.3 141.1 167.0

lowest :   0.00  24.46  24.51  24.52  59.00, highest: 194.00 194.90 195.00 200.00 200.10
```

agua



```
   n  missing  distinct   Info   Mean    Gmd    .05    .10    .25    .50    .75    .90    .95
1030        0       205  0.998  181.6  23.82  146.1  154.6  164.9  185.0  192.0  203.5  228.0

lowest : 121.75 126.60 127.00 127.30 137.80, highest: 228.00 236.70 237.00 246.90 247.00
```

super__plastificante



```
       n  missing  distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95 
    1030        0       155     0.95    6.203    6.426     0.00     0.00     0.00     6.35    10.16    12.21    16.05 

lowest :  0.00  1.72  1.90  2.00  2.20, highest: 22.00 22.10 23.40 28.20 32.20
```

---

agregador__grosso

```
       n  missing  distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95 
    1030        0       284        1    972.9    88.55    842.0    852.1    932.0    968.0   1029.4   1076.5   1104.0 

lowest :  801.0  801.1  801.4  811.0  814.0, highest: 1124.4 1125.0 1130.0 1134.3 1145.0
```

---

agregador__fino

```
       n  missing  distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95 
    1030        0       304        1    773.6    89.87    613.0    664.1    730.9    779.5    824.0    880.8    898.1 

lowest : 594.0 605.0 611.8 612.0 613.0, highest: 925.7 942.0 943.1 945.0 992.6
```

---

idade

```
       n  missing  distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95 
    1030        0        14    0.925    45.66    50.89        3        3        7       28       56      100      180 

lowest :   1   3   7  14  28, highest: 120 180 270 360 365
                                                                                          
Value          1     3     7    14    28    56    90    91   100   120   180   270   360   365
Frequency      2   134   126    62   425    91    54    22    52     3    26    13     6    14
Proportion 0.002 0.130 0.122 0.060 0.413 0.088 0.052 0.021 0.050 0.003 0.025 0.013 0.006 0.014
```

---

forca__compressiva

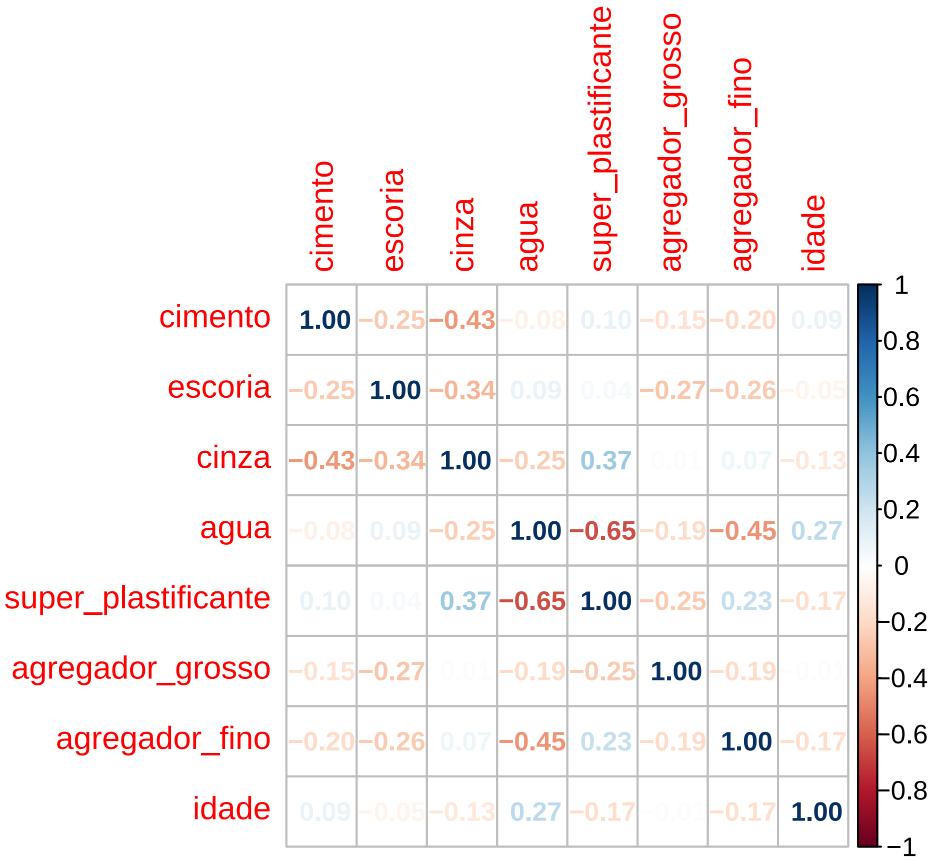
```
       n  missing  distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95 
    1030        0       938        1    35.82    18.92    10.96    14.20    23.71    34.44    46.14    58.82    66.80 

lowest :  2.331808  3.319827  4.565021  4.782206  4.827711
highest: 79.400056 79.986111 80.199848 81.751169 82.599225
```

Preparação dos dados

```r
## Separando o conjunto de dados em treino e teste
set.seed(2)
inTrain <- createDataPartition(dados$forca_compressiva, p = 7/10)[[1]]
treino <- dados[inTrain,]
teste <- dados[-inTrain,]

## Mantendo casos completos em treino e teste
treino <- treino[complete.cases(treino),]
teste <- teste[complete.cases(teste),]

## Separando a variavel resposta, categóricas e numericas
resposta <- treino$forca_compressiva
resposta_teste <- teste$forca_compressiva

## Removendo a variável resposta
treino <- treino[,-ncol(treino)]
teste <- teste[,-ncol(teste)]

## Retendo as numéricas
Ind_numericas <- colnames(treino)[sapply(treino, is.numeric)]
Ind_categoricas <- colnames(treino)[sapply(treino, function(x) !is.numeric(x))]
numericas <- treino[,Ind_numericas]
categorias <- treino[,Ind_categoricas]
```

Redução de dimensionalidade

```
## Analisando as correlações
M <- cor(numericas, use = 'complete.obs')
corrplot(M, method='number', diag = T, number.cex = 0.8)
```



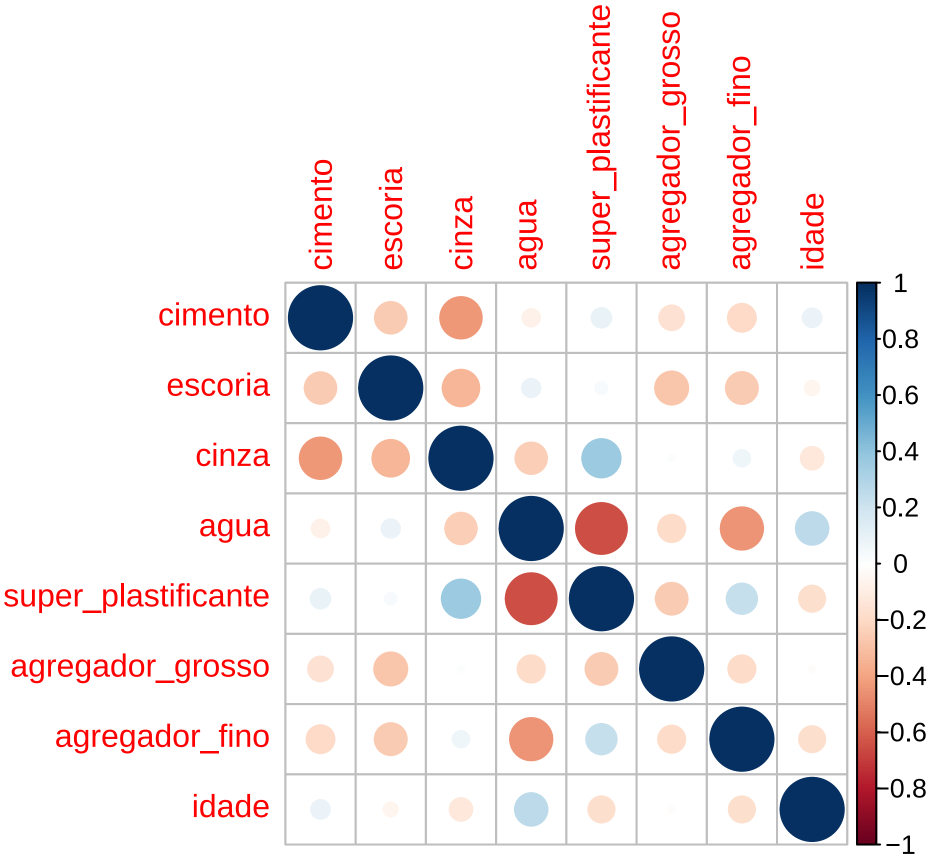
```
summary(M[upper.tri(M)])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
## -0.64810 -0.25116 -0.16258 -0.11522  0.04417  0.36742
```

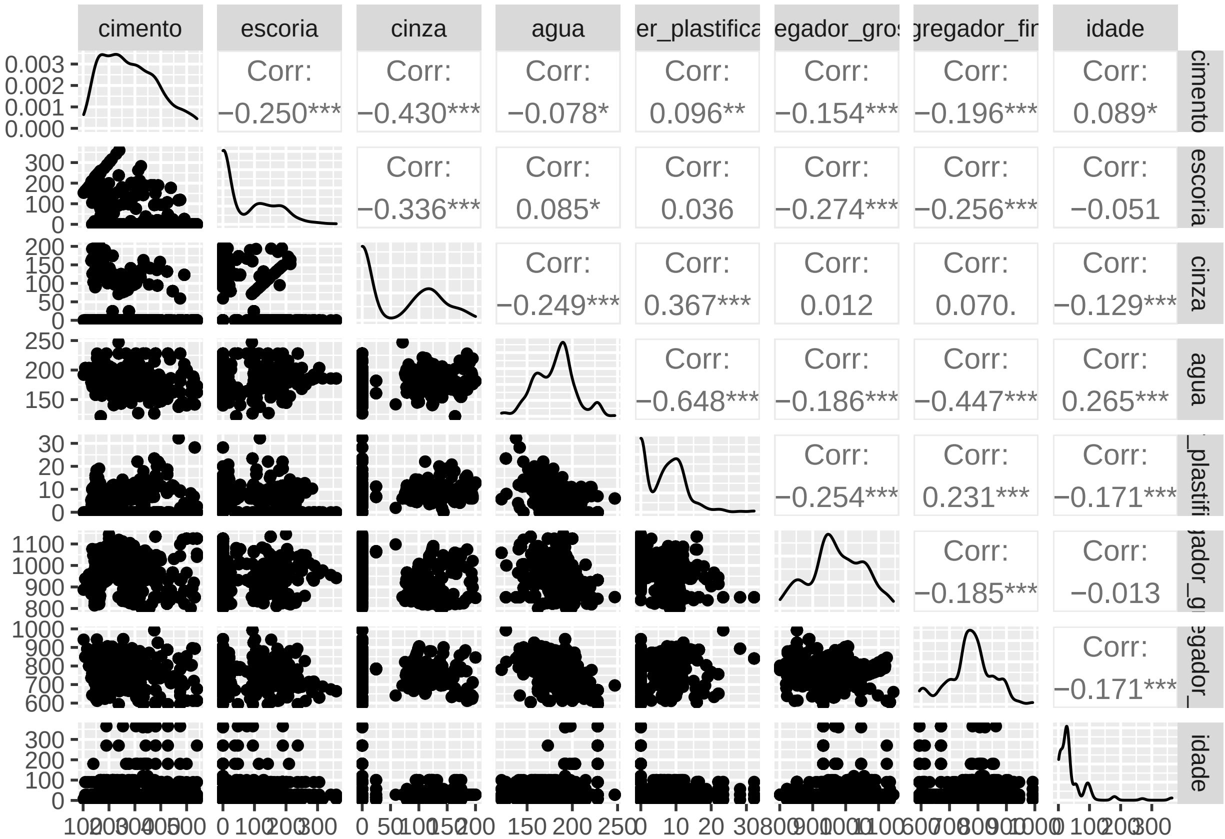
```
## Imprimindo as correlações na forma de circulos
M <- cor(numericas, use = 'complete.obs')
summary(M[upper.tri(M)])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
## -0.64810 -0.25116 -0.16258 -0.11522  0.04417  0.36742
```

```
corrplot(M, method='circle')
```



```
## Visualizando as correlações
ggpairs(numericas)
```


cimento
escoria
cinza
agua
er_plastifica
egador_gros
gregador_fir
idade
Corr:
−0.250***
Corr:
−0.430***
Corr:
−0.078*
Corr:
0.096**
Corr:
−0.154***
Corr:
−0.196***
Corr:
0.089*
Corr:
−0.336***
Corr:
0.085*
Corr:
0.036
Corr:
−0.274***
Corr:
−0.256***
Corr:
−0.051
Corr:
−0.249***
Corr:
0.367***
Corr:
0.012
Corr:
0.070.
Corr:
−0.129***
Corr:
−0.648***
Corr:
−0.186***
Corr:
−0.447***
Corr:
0.265***
Corr:
−0.254***
Corr:
0.231***
Corr:
−0.171***
Corr:
−0.185***
Corr:
−0.013
Corr:
−0.171***
cimento
escoria
cinza
agua
_plastifi
jador_g
egador_
idade
0.003
0.002
0.001
0.000
300
200
100
0
200
150
100
50
0
250
200
150
30
20
10
0
1100
1000
900
800
1000
900
800
700
600
300
200
100
0
100200300400500
0 100 200 300
0 50 100 150 200
150 200 250
0 10 20 30
800 900 1000 1100
600 700 800 900 1000
0 100 200 300

Modelagem