

Análise dos auditores das Cooperativas de Crédito em 2021

Ricardo Theodoro

2022-10-18

Objetivo

O objetivo deste *script* é analisar os dados que envolvem os auditores independentes das cooperativas de crédito brasileiras em 2021. Para isso, iremos utilizar os dados disponibilizados pelo Banco Central do Brasil (BACEN), contendo diversas informações sobre as características destas cooperativas.

Nesta base de dados estão apenas as cooperativas de crédito que informaram seus auditores independentes ao BACEN em 2021, as que não informaram foram retiradas.

Os pacotes utilizados para alcançarmos estes objetivos serão:

```
# eval = FALSE mostra o código, mas não roda

# Pacotes

install.packages("dplyr")
install.packages("glue")
install.packages("janitor")
```

Importação de dados

O primeiro passo para o tratamento da base é importar os dados, para isso vamos utilizar o *chunk* abaixo.

```
# Importando a base
coop_cred_2021_auditores <- read.csv("data_raw/coop_cred_2021_auditores.csv") |>
  janitor::clean_names()
```

Como é possível observar, nossa base de dados possui 35 variáveis, sendo elas: **cnpj**, **nome_coop**, **auditor_independente**, **mudou_auditor**, **situacao**, **uf**, **endereco_eletronico**, **segmento_prudencial**, **numero_agencias**, **ativo_total**, **passivo_total**, **patrimonio_liquido**, **despesas_operacionais**, **receitas_operacionais**, **depositos_totais**, **caixa**, **disponibilidades**, **sobras**, **peclid**, **capital_social**, **compensacao**, **carteira_credito**, **classificacao_carteira_credito**, **regiao**, **municipio**, **classe**, **criterio_de_associacao**, **categ_coop_sing**, **filiacao**, **big_four**, **data_inicio_atividades**, **idade_em_2022**, **total_de_cooperados**, **cooperados_pf** e **cooperados_pj**.

Agora, vamos ver um resumo dos dados importados para escolhermos quais variáveis iremos utilizar e quais precisaremos tratar.

```
dplyr::glimpse(coop_cred_2021_auditores)

## Rows: 392
## Columns: 35
## $ cnpj                <int> 5790149, 309024, 1634601, 10398952, 315~
## $ nome_coop            <chr> "CENTRAL COOPERATIVA DE CRÉDITO NO ESTA~
## $ auditor_independente <chr> "MOORE PRISMA AUDITORES INDEPENDENTES",~
## $ mudou_auditor        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
```

```
## $ situacao      <chr> "Autorizada em Atividade", "Autorizada ~
## $ uf           <chr> "ES", "MG", "RS", "SC", "SP", "SP", "RS~
## $ endereco_eletronico <chr> "www.cecoop.com.br", "www.cecremge.org.~
## $ segmento_prudencial <chr> "S4", "S4", "S4", "S4", "S4", "S4", "S5~
## $ numero_agencias <int> 0, 0, 0, 0, 1, 8, 5, 0, 0, 0, 0, 1, ~
## $ ativo_total   <dbl> 577843270, 972721594887, 211132462184, ~
## $ passivo_total <dbl> 577843270, 972721594887, 211132462184, ~
## $ patrimonio_liquido <dbl> 106841099, 28976465231, 10205121563, 18~
## $ despesas_operacionais <dbl> -144941515, -7964209117, -3386032690, --
## $ receitas_operacionais <dbl> 140941515, 8250858860, 3405587104, 4511~
## $ depositos_totais <dbl> 363017390, 40136012, NA, NA, NA, 639729~
## $ caixa         <chr> "351,70", NA, "7706,46", "49,60", "6977~
## $ disponibilidades <dbl> 8014555, 300000, 1070646, 9819688, 7053~
## $ sobras        <dbl> NA, 327608050, NA, 40287195, 739196642,~
## $ pecld         <dbl> NA, -228434660, -2151399, -520000, NA, ~
## $ capital_social <dbl> 106841099, 25698125393, 9998573090, 174~
## $ compensacao   <dbl> 1882904, 220180962654, 36300642623, 331~
## $ carteira_credito <dbl> NA, 34850267266, 430279801, 104000000, ~
## $ classificacao_carteira_credito <dbl> NA, 34850267266, 430279801, 104000000, ~
## $ regio         <chr> "Sudeste", "Sudeste", "Sul", "Sul", "Su~
## $ municipio     <chr> "VITORIA", "BELO HORIZONTE", "PORTO ALE~
## $ classe        <chr> "Central", "Central", "Central", "Confe~
## $ criterio_de_associacao <chr> NA, NA, NA, NA, NA, "Livre Admissão", "~
## $ categ_coop_sing <chr> NA, NA, NA, NA, NA, "Plena", "Clássica"~
## $ filiacao       <chr> NA, NA, "CONF NAC COOP CENTRAIS UNICRED~
## $ big_four      <int> 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, ~
## $ data_inicio_atividades <chr> "2003-07-17", "1994-11-22", "1996-09-14~
## $ idade_em_2022 <int> 19, 28, 26, 14, 28, 46, 38, 8, 20, 26, ~
## $ total_de_cooperados <int> NA, NA, NA, NA, NA, 8665, 2198, NA, NA,~
## $ cooperados_pf <int> NA, NA, NA, NA, NA, 6485, 2195, NA, NA,~
## $ cooperados_pj <int> NA, NA, NA, NA, NA, 2180, 3, NA, NA, NA~
```

É possível observar que temos tanto variáveis categóricas (chr) quanto numéricas (dbl e int). Será preciso verificar se todas estão no formato correto. E, como nós não iremos utilizar todos os dados, então vamos selecionar apenas as colunas de interesse, que são:

```
coop_cred_2021_audidores <- coop_cred_2021_audidores |>
  dplyr::select(cnpj, auditor_independente, big_four, uf, numero_agencias,
               ativo_total, passivo_total, filiacao)
```

Agora, vamos analisar as estatísticas descritivas das variáveis numéricas e verificar se estão no formato correto para uso:

```
library(tidyselect)

coop_cred_2021_audidores |>
  dplyr::select(where(is.double), -cnpj) |>
  summary()
```

```
##   ativo_total      passivo_total
## Min.   :    60127828 Min.   :    60127828
## 1st Qu.:  2042443547 1st Qu.:  2042443547
## Median : 12756650864 Median : 12756650864
## Mean   :  95752784862 Mean   :  95752784862
## 3rd Qu.:  73512729828 3rd Qu.:  73512729828
## Max.   :4445450799630 Max.   :4445450799630
```

```
# A função `where` não funciona no formato
# tidyselect::where(), por isso usamos o library()

# Removemos o cnpj pois ele serve apenas para identificação
```

Aqui, podemos observar que nem todas as variáveis que contêm números estão classificadas como numéricas, para isso precisamos transformá-las

```
# Primeira forma
coop_cred_2021_audidores$big_four <- as.numeric(coop_cred_2021_audidores$big_four)

# Segunda forma

coop_cred_2021_audidores <- coop_cred_2021_audidores |>
  dplyr::mutate(
    numero_agencias = as.numeric(numero_agencias )
  )
```

Agora, rodamos novamente o código para obtermos as estatísticas descritivas das variáveis numéricas

```
coop_cred_2021_audidores |>
  dplyr::select(where(is.double), -cnpj) |> # aqui removemos o cnpj pois ele serve apenas para identif
  summary()
```

```
##      big_four      numero_agencias      ativo_total
## Min.      :0.000   Min.       : 0.00   Min.       :    60127828
## 1st Qu.:0.000   1st Qu.: 0.00   1st Qu.:   2042443547
## Median :0.000   Median : 1.00   Median :  12756650864
## Mean   :0.293   Mean   : 5.72   Mean   :  95752784862
## 3rd Qu.:1.000   3rd Qu.: 7.00   3rd Qu.:  73512729828
## Max.    :1.000   Max.    :118.00  Max.     :4445450799630
## passivo_total
## Min.      :    60127828
## 1st Qu.:   2042443547
## Median :  12756650864
## Mean   :  95752784862
## 3rd Qu.:  73512729828
## Max.    :4445450799630
```

Exploração de dados

Análise gráfica