

Análise dos auditores das Cooperativas de Crédito em 2021

Ricardo Theodoro

19 de outubro de 2022

Sumário

Introdução	1
Objetivo	1
Metodologia	1
Importação e tratamento de dados	1
Correlação entre variáveis	6
Análise gráfica	8

Introdução

Aqui iremos aprender a criar um relatório automatizado para análise de dados.

Além do que veremos neste *script*, você pode continuar estudando em R Markdown Cookbook.

Neste *script* as análises e tratamentos não serão completas e detalhadas, serão apenas exemplos para que você possa criar seus relatórios utilizando os conhecimentos adquiridos nas aulas passadas e, também, no decorrer desta.

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.

Metodologia

Os pacotes utilizados estão listados no *chunk*¹ abaixo. Esse *chunk* não irá o código, apenas exibir seu conteúdo. Caso você não tenha algum destes pacotes instalado, estas linhas devem ser rodadas manualmente, antes de gerar o relatório.

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

# Pacotes
install.packages("dplyr")
install.packages("glue")
install.packages("janitor")
install.packages("kableExtra")
install.packages("tidyselect")
```

Importação e tratamento de dados

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

¹ Trecho de código que pode ou não ser exibido no relatório, assim como pode estar ativo ou não.

```
# Importando a base
coop_cred_2021_audidores <-
  read.csv("data_raw/coop_cred_2021_audidores.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**, **endereço_eletronico**, **segmento_prudencial**, **numero_agencias**, **ativo_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**, **cooperados_pj** e **passivo_total**.

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

```
# comment = NA remove as '#' quando exibe os resultados

dplyr::glimpse(coop_cred_2021_audidores)
```

```
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.cecocoop.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, 0, 1, ~
$ ativo_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,~
$ peclid              <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, ~
$ regiao              <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~
$ passivo_total       <dbl> 577843270, 972721594887, 211132462184, ~
```

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, patrimonio_liquido,
               filiacao)
```

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

```
coop_cred_2021_audidores |>
  dplyr::select(-cnpj) |>
  stargazer::stargazer(
    title = "Estatísticas Descritivas das variáveis numéricas - pré-tratamento",
    summary = TRUE,
    iqr = TRUE,
    flip = TRUE,
    header = FALSE
  )
```

Tabela 1: Estatísticas Descritivas das variáveis numéricas - pré-tratamento

Statistic	big_four	numero_agencias	ativo_total	patrimonio_liquido
N	392	392	392	392
Mean	0.293	5.722	95,752,784,861.000	5,403,911,723.000
St. Dev.	0.456	12.190	307,742,625,442.000	15,136,438,783.000
Min	0	0	60,127,828	14,682,260
Pctl(25)	0	0	2,042,443,547	466,006,482.0
Pctl(75)	1	7	73,512,729,828	4,253,064,882.0
Max	1	118	4,445,450,799,632	171,822,871,461

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

Aqui, podemos observar que nem todas as variáveis que contém 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(-cnpj) |>
  stargazer::stargazer(
    title = "Estatísticas Descritivas das Variáveis Numéricas",
    summary = TRUE,
    iqr = TRUE,
    flip = TRUE,
    header = FALSE
  )
```

Agora iremos analisar as variáveis categóricas, da seguinte forma:

Tabela 2: Estatísticas Descritivas das Variáveis Numéricas

Statistic	big_four	numero_agencias	ativo_total	patrimonio_liquido
N	392	392	392	392
Mean	0.293	5.722	95,752,784,861.000	5,403,911,723.000
St. Dev.	0.456	12.190	307,742,625,442.000	15,136,438,783.000
Min	0	0	60,127,828	14,682,260
Pctl(25)	0	0	2,042,443,547	466,006,482.0
Pctl(75)	1	7	73,512,729,828	4,253,064,882.0
Max	1	118	4,445,450,799,632	171,822,871,461

```
# Função
resumo_categorica <- function(data, var) {
  teste <- tidyselect::enquo(var) |> rlang::as_name()

  data |>
    dplyr::group_by({{ var }}) |>
    dplyr::count() |>
    dplyr::arrange(-n) |>
    janitor::adorn_totals() |>
    kableExtra::kable(format = "pipe",
                      caption = glue::glue("Informações de {teste}"))
}

resumo_categorica(coop_cred_2021_auditores, auditor_independente)
```

Tabela 3: Informações de auditor_independente

auditor_independente	n
ERNST & YOUNG AUDITORES INDEPENDENTES S/S	67
MOORE PRISMA AUDITORES INDEPENDENTES	52
PRICEWATERHOUSECOOPERS AUDITORES INDEPENDENTES	39
PADRAO AUDITORIA S/S	37
BAUER AUDITORES ASSOCIADOS	28
MANDARINO & ASSOCIADOS AUDITORES	24
SACHO - AUDITORES INDEPENDENTES	17
CONTROL AUDITORIA E CONTABILIDADE	15
D'AGOSTINI CONSULTORIA E AUDITORIA S/S	15
ANEND AUDITORES INDEPENDENTES	11
LINEAR AUDITORES INDEPENDENTES SOCIEDADE SIMPLES	9
RUSSELL BEDFORD GM AUDITORES INDEPENDENTES S/S	9
MBAUDIT - AUDITORES INDEPENDENTES S/S	7
IDERALDO LUIZ AGOSTINHO DE MORAES	6
PWC AUDITORES INDEPENDENTES	5
COOPERAUDI AUDITORES INDEPENDENTES	4
FABRICIO KOEKE	3
HILDO JARDIM ALEGRIA	3
JOAO CARLOS SUBACZ	3
RAMIRES E CIA - AUDITORES INDEPENDENTES SOCIEDADE SIMPLES	3
APPROACH AUDITORES INDEPENDENTES S/S	2
AUDICONSULT AUDITORES SS	2
HELIO MAZZI JUNIOR	2
PAULO SERGIO SANGIORGIO	2
ROKEMBACH + LAHM, VILLANOVA & CIA AUDITORES	2

auditor_independente	n
ADIRLEY GASPARIN	1
ANTONIO LUCIO PEREIRA SANTOS	1
AUDSERVICE-AUDITORES ASSOCIADOS S/C	1
CHOSUKE KOEKE	1
COKINOS & ASSOCIADOS AUDITORES INDEPENDENTES	1
CONFEDERACAO NACIONAL DE AUDITORIA COOPERATIVA - CNAC	1
EDIMAR WANDERLEY	1
ERNST & YOUNG A I S/S	1
Ernst and Young	1
ERNST AND YOUNG	1
FABIO EDUARDO DE ALMEIDA BAUER	1
FRANCISCO ASSIS DE SOUSA	1
GRUNITZKY - AUDITORES INDEPENDENTES S/S	1
HILDEBRANDO CAMARGO	1
HUGO FRANCISCO SACHO	1
IVANILDO ALVES MESSIAS	1
JPS-AUDITORES & CONSULTORES	1
KPMG AUDITORES INDEPENDENTES	1
LAURO ANGELO CERUTTI	1
MARCELINO VASCONCELOS BARROSO	1
MAURILIO ROSA	1
MAZARS AUDITORES INDEPENDENTES - SOCIEDADE SIMPLES	1
PEPPE ASSOCIADOS - CONSULTORES & AUDITORES INDEPENDENTES	1
SERGIO NOBORU OUTAKA	1
VAZ & MAIA AUDITORES INDEPENDENTES	1
Total	392

```
resumo_categorica(coop_cred_2021_audidores, filiacao)
```

Tabela 4: Informações de filiacao

filiacao	n
NA	179
CCCR CRESOL CENTRAL BRASIL	28
CCC ESTADO SP - CECRESP	20
CCC INT SOL CENTRAL CRESOL BASER	19
CCCR CRESOL SICOPER	19
CENTRAL SICREDI NORTE/NORDESTE	18
COOPCENTRAL AILOS	13
CCCM UNICRED CENTRAL RS	12
CC UNICRED CENTRAL MULTIRREGIONAL LTDA. - UCM	11
CCC MINAS GERAIS - CREDIMINAS	10
CREDISIS CENTRAL DE COOPERATIVAS DE CRÉDITO LTDA.	8
CCC DO ESTADO DE SÃO PAULO	7
CCC DOS EST DE MT, MS E MUN DE CACOAL/RO	7
CCC UNICRED CENTRAL CONEXÃO LTDA - UNICRED CENTRAL CONEXÃO	6
UNICRED CENTRAL RJ/MT	5
UNIPRIME CENTRAL CCC LTDA.	5
CCCM DO RS, SC E PR	4
CONF NAC COOP CENTRAIS UNICRED	4
CRESOL CONFEDERAÇÃO	4
CC ASCOOB CENTRAL	3
CENTRAL COOPERATIVA DE CRÉDITO NO ESTADO DO ESPÍRITO SANTO	3

filiacao	n
CCCM SICOOB UNIMAIS	2
CENTRAL SICOOB UNI DE CC	2
CCC DE GOIAS LTDA	1
CCC DO NORTE DO BRASIL	1
SICOOB CENTRAL NE	1
Total	392

```
resumo_categorica(coop_cred_2021_audidores, uf)
```

Tabela 5: Informações de uf

uf	n
SP	100
SC	57
RS	53
MG	38
PR	31
RJ	31
ES	13
BA	10
RO	10
MT	7
PB	7
GO	6
AL	4
MS	4
PA	4
PE	4
CE	3
DF	2
MA	2
RN	2
SE	2
AC	1
PI	1
Total	392

Tabela 6: Proporção de Cooperativas de Crédito Auditadas por Big Four

big_four	n	percent
0	277	0.7066
1	115	0.2934

Como podemos observar na tabela acima, das 392 cooperativas da amostra, apenas 115 são auditadas por uma empresa considerada *Big Four*.

Correlação entre variáveis

Nesta parte iremos testar se existe ou não correlação entre as variáveis que iremos utilizar. Iremos testar tanto as variáveis qualitativas (categóricas) quanto as quantitativas (numéricas), e as duas juntas. Pra isso iremos utilizar os testes χ^2 , **Pearson**, R^2 , **covariância** e **correlação**.

- Associação entre variáveis qualitativas (χ^2 e Pearson)

```
chisq_filiacao <- chisq.test(
  coop_cred_2021_auditores$big_four,
  coop_cred_2021_auditores$filiacao
)
```

```
## Warning in chisq.test(coop_cred_2021_auditores$big_four,
## coop_cred_2021_auditores$filiacao): Aproximação do qui-quadrado pode estar
## incorreta
```

```
print(chisq_filiacao)
```

```
##
## Pearson's Chi-squared test
##
## data: coop_cred_2021_auditores$big_four and coop_cred_2021_auditores$filiacao
## X-squared = 186, df = 24, p-value <2e-16
```

```
chisq_uf <- chisq.test(
  coop_cred_2021_auditores$big_four,
  coop_cred_2021_auditores$uf
)
```

```
## Warning in chisq.test(coop_cred_2021_auditores$big_four,
## coop_cred_2021_auditores$uf): Aproximação do qui-quadrado pode estar incorreta
```

```
print(chisq_uf)
```

```
##
## Pearson's Chi-squared test
##
## data: coop_cred_2021_auditores$big_four and coop_cred_2021_auditores$uf
## X-squared = 116, df = 22, p-value = 9e-15
```

Como podemos observar nos resultados dos testes χ^2 acima, o temos que os p-valores são de 6.0273×10^{-27} , e 8.9574×10^{-15} . Ou seja, menores que 0,05. Isso indica que existem diferenças de filiação e UF entre os grupos, auditados por *Big Four* e auditados por outras empresas.

Agora, para verificar se existem correlações entre as variáveis, temos que realizar o teste de *Pearson*.

TESTAR GoodnessOfFit - TESTE DE CORRELAÇÃO

- Associação entre variáveis quantitativas (covariância e correlação)

```
# Diferença entre grupos
```

```
chisq.test(
  coop_cred_2021_auditores$big_four,
  coop_cred_2021_auditores$patrimonio_liquido
)
```

```
## Warning in chisq.test(coop_cred_2021_auditores$big_four,
## coop_cred_2021_auditores$patrimonio_liquido): Aproximação do qui-quadrado pode
## estar incorreta
```

```
##
## Pearson's Chi-squared test
##
## data: coop_cred_2021_auditores$big_four and coop_cred_2021_auditores$patrimonio_liquido
## X-squared = 392, df = 391, p-value = 0.5
```

```
chisq.test(
  coop_cred_2021_auditores$big_four,
  coop_cred_2021_auditores$ativo_total
)
```

```
## Warning in chisq.test(coop_cred_2021_auditores$big_four,
## coop_cred_2021_auditores$ativo_total): Aproximação do qui-quadrado pode estar
## incorreta

##
## Pearson's Chi-squared test
##
## data: coop_cred_2021_auditores$big_four and coop_cred_2021_auditores$ativo_total
## X-squared = 392, df = 391, p-value = 0.5
```

Neste dois casos, como o p-valor de ambos não são menores que 0,05, significa que as distribuições de ambos os grupos são iguais.

```
# Correlação
cor(
  coop_cred_2021_auditores |> dplyr::filter(big_four == 0) |> dplyr::select(patrimonio_liquido),
  coop_cred_2021_auditores |> dplyr::filter(big_four == 0) |> dplyr::select(ativo_total),
  method = "pearson"
)
```

```
##                ativo_total
## patrimonio_liquido    0.7614
```

```
cor(
  coop_cred_2021_auditores |> dplyr::filter(big_four == 1) |> dplyr::select(patrimonio_liquido),
  coop_cred_2021_auditores |> dplyr::filter(big_four == 1) |> dplyr::select(ativo_total),
  method = "pearson"
)
```

```
##                ativo_total
## patrimonio_liquido    0.8815
```

Quanto mais próximo de 1, melhor. Neste caso, podemos observar que em ambos os grupos, quanto maior é o Ativo Total, maior é o Patrimônio Líquido.

```
# Covariância
cov(
  coop_cred_2021_auditores$patrimonio_liquido,
  coop_cred_2021_auditores$ativo_total
)
```

```
## [1] 4.105e+21
```

Alta covariância.

- Associação entre variáveis qualitativas e quantitativas (R^2)

```
# R2
r2 <- lm(patrimonio_liquido ~ big_four, coop_cred_2021_auditores) |> summary()
r2$r.squared
```

```
## [1] 0.08463
```

Quanto mais próximo de 1, melhor. Neste caso, vemos que o R^2 é de 0.0846, o que significa que o fato da cooperativa de crédito ser auditada por uma *Big Four* não tem relação com seu tamanho.

Análise gráfica

```
coop_cred_2021_auditores |>
  ggplot2::ggplot() +
  ggplot2::aes(x = patrimonio_liquido, y = big_four) +
  ggplot2::geom_point(colour = "#011e5a") +
  ggplot2::xlab("Patrimônio Líquido") +
  ggplot2::ylab("Big Four") +
  ggplot2::geom_smooth(method = "lm", se = FALSE)
```

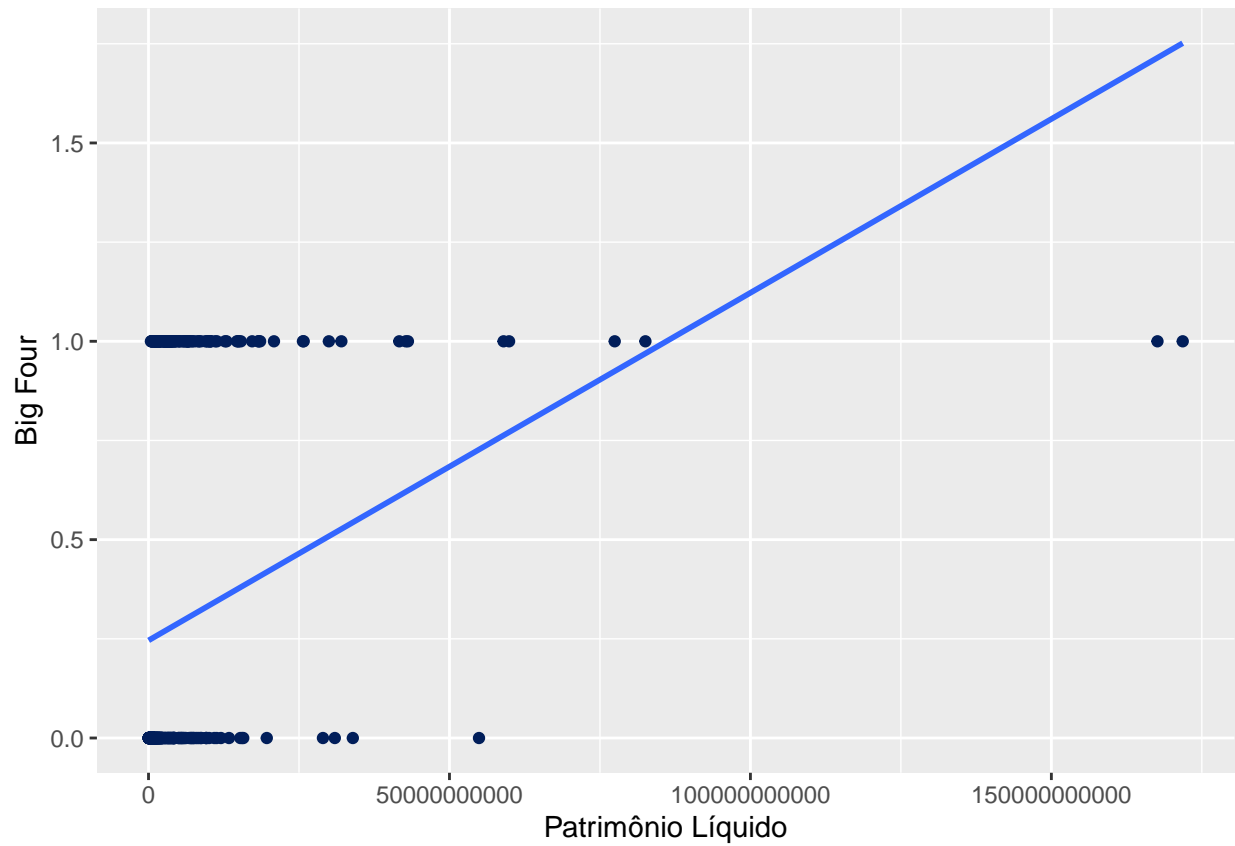


Figura 1: Relação PL x Big Four