

# Identificando cooperativas de crédito auditadas por Big Four

## Procedimentos metodológicos

Ricardo Theodoro nUSP 10.191.222

### Introdução

Uma cooperativa, na definição de Barton (1983), é uma empresa pertencente e controlada pelo usuário que distribui benefícios com base no uso. Fundamentado nesse conceito, são extraídos três aspectos característicos desta organização: primeiro, as pessoas que possuem e financiam a cooperativa são aqueles que a utilizam; segundo, o controle da cooperativa é feito pela escolha de quem utiliza a cooperativa; terceiro, os benefícios oferecidos para o usuário, revelam-se como negócio a custo.

Essa característica de controle pelos usuários se relacionam com o princípio da gestão democrática, garantindo que o controle da cooperativa seja dos seus próprios membros, sendo que para essa forma de gestão ocorrer, os associados elegem representantes, também associados, que os representam na administração da sociedade (Schaefer, Bittencourt, e Ferraz 2022). Assim, verifica-se uma relação de agência, compreendida como um contrato sob o qual uma ou mais pessoas, definidas como principais, emprega outra pessoa (agente), para executar em seu nome um serviço que implique a delegação de algum poder de decisão do agente (Jensen e Meckling 1976).

Nesta relação da agência, os interesses do principal e do agente podem divergir, assim no intuito de manter os interesses alinhados, as partes irão incorrer em custos de agência, que são a soma: das despesas de monitoramento por parte do principal; das despesas com a concessão de garantias contratuais por parte do agente; e do custo residual (Jensen e Meckling 1976). Dentre as despesas de monitoramento a serem assumidas pelo principal, está a auditoria independente, que tem como atribuição básica verificar se as demonstrações financeiras refletem, em todos os aspectos relevantes, a posição econômica e financeira da cooperativa, assim como a revisão e avaliação dos controles internos (Schaefer, Bittencourt, e Ferraz 2022). Ainda, concentrando o mercado de atuação da auditoria independente, tem-se as chamadas *Big Four* (Deloitte; Ernst & Young; KPMG e PwC).

Sabe-se que, a efetividade da auditoria como um instrumento de monitoramento é comprovada por estudo, como Le e Lobo (2020), que ao documentarem o papel que a auditoria desempenha na conformidade das demonstrações financeiras.

Desse modo, busca-se responder a questão: quais características das cooperativas de crédito brasileiras influenciam na opção de contratarem uma *Big Four*?

Como justificativa, tem-se a incipiente discussão sobre o mercado de auditoria independente atuante em cooperativas de crédito, sobretudo na identificação de concentração do mercado em *Big Four*, que tem merecido cada vez mais a atenção, não permitindo mais negligenciá-lo. Ainda, a escolha das cooperativas de crédito está relacionada ao seu grau de importância em promover o crescimento, desenvolvimento econômico e social, sendo as maiores instituições financeiras de varejo no Brasil.

### Procedimentos metodológicos

#### Base de dados

Foram selecionadas apenas as cooperativas de crédito singulares que informaram quem eram seus auditores independentes ao Banco Central do Brasil (BACEN) em outubro de 2021, resultando em uma amostra de 365 instituições. Assim, na sequência foram selecionadas contas vindas do balanço contábil que em estudos anteriores

foram utilizadas para identificar gerenciamento de resultados, que são: Receitas Operacionais (ROP); Despesas Operacionais (DES); e Perdas Estimadas em Créditos de Liquidação Duvidosa (PECLD). A escolha foi baseada na literatura existente sobre instituições financeiras e em empresas do terceiro setor, uma vez que são passíveis de alterações devido a escolhas do gestor (Theodoro, Bonacim, e Moura Costa 2021).

Também foram selecionadas contas que podem sinalizar o tamanho da cooperativa como Ativo Total e Patrimônio Líquido, visto que podemos pensar que quanto maior a cooperativa, maior sua preocupação com a qualidade dos dados auditados. Além destas, também foi selecionado o número de agências e quantidade de cooperados.

Salienta-se que os dados das contas se referem ao mês de dezembro de 2020, uma vez que o balanço consolidado de dezembro de 2021 ainda não foi disponibilizado pelo BACEN. Ainda foram adicionados dados cadastrais sobre Classe, Critério de Associação e Filiação e Idade, que foi calculada através da data de início de atividade conforme cadastro do CNPJ junto à Receita Federal Brasileira em outubro de 2021, para futura melhoria no modelo.

```
# Tratamento da base
coop <- read.csv("data/auditorIndependente_coopcred.csv") |>
  dplyr::filter(data_coleta == "10/2021" & stringr::str_trim(classe) == "Singular") |>
  janitor::clean_names() |>
  dplyr::mutate(
    ativo_total = stringr::str_replace(ativo_total, ",", "."),
    ativo_total = as.numeric(ativo_total),
    patrimonio_liquido = stringr::str_replace(patrimonio_liquido, ",", "."),
    patrimonio_liquido = as.numeric(patrimonio_liquido),
    sobras = stringr::str_replace(sobras, ",", "."),
    sobras = as.numeric(sobras),
    pecld = stringr::str_replace(pecld, ",", "."),
    pecld = as.numeric(pecld),
    receitas_operacionais = stringr::str_replace(receitas_operacionais, ",", "."),
    receitas_operacionais = as.numeric(receitas_operacionais),
    despesas_operacionais = stringr::str_replace(despesas_operacionais, ",", "."),
    despesas_operacionais = as.numeric(despesas_operacionais)
  ) |>
  dplyr::select(
    -data_coleta, -mudou_auditor, -situacao, -uf, -endereco_eletronico,
    -regiao, -municipio, -classe
  )

coop[is.na(coop)] <- 0
```

Com isso, temos a seguinte estatística descritiva para a amostra total selecionada:

```
library(magrittr)

# Estatística descritiva
coop |>
  dplyr::select(
    idade_em_2022,
    numero_agencias,
    total_de_cooperados,
    ativo_total,
    patrimonio_liquido,
    despesas_operacionais,
    receitas_operacionais,
    sobras,
```

```

  peclid
) %>%
stargazer::stargazer(
  ... = .,
  type = "text",
  title = "Estatística Descritiva da amostra completa de cooperativas de crédito singulares",
  style = "aer",
  decimal.mark = ".",
  digits = 0
)

```

Estatística Descritiva da amostra completa de cooperativas de crédito singulares

Statistic	N	Mean	St. Dev.	Min	Max
idade_em_2022	365	28	11	8	56
numero_agencias	365	6	13	0	118
total_de_cooperados	365	10,691	44,272	0	794,684
ativo_total	365	75,883,409,033	291,107,553,122	60,127,828	4,445,450,799,632
patrimonio_liquido	365	4,760,432,371	14,951,146,194	14,682,260	171,822,871,461
despesas_operacionais	365	-1,177,708,654	3,520,938,908	-42,527,929,368	-2,270,470
receitas_operacionais	365	1,461,134,105	4,383,848,517	1,873,083	53,567,939,894
sobras	365	248,235,877	1,112,247,323	-708,226,808	14,858,323,408
peclid	365	-669,858,856	2,576,867,057	-28,956,184,003	-21,920

Com estes dados é possível observar que a amostra é composta por cooperativas de crédito com características distintas quanto ao seu tamanho, com alto desvio padrão, tendo cooperativas desde muito pequenas até muito grandes para todas as variáveis.

Ainda é possível comparar as estatísticas descritivas separadas pelos grupos, sendo 0 auditadas por empresa comum e 1 auditada por *Big Four*.

```

coop |>
dplyr::filter(big_four == 0) |>
dplyr::select(
  idade_em_2022,
  numero_agencias,
  total_de_cooperados,
  ativo_total,
  patrimonio_liquido,
  despesas_operacionais,
  receitas_operacionais,
  sobras,
  peclid
) %>%
stargazer::stargazer(
  ... = .,
  type = "text",
  title = "Estatística Descritiva da amostra de cooperativas de crédito singulares não auditadas por Big F",
  style = "aer",
  decimal.mark = ".",

```

```
digits = 0
)
```

Estatística Descritiva da amostra de cooperativas de crédito singulares não auditadas por Big Four

Statistic	N	Mean	St. Dev.	Min	Max
idade_em_2022	266	30	12	8	56
numero_agencias	266	4	8	0	70
total_de_cooperados	266	5,394	9,477	0	81,561
ativo_total	266	32,819,741,374	87,210,631,215	60,127,828	893,470,031,934
patrimonio_liquido	266	2,415,536,049	5,137,815,749	14,682,260	54,907,133,319
despesas_operacionais	266	-511,859,161	1,034,198,658	-10,575,329,401	-2,270,470
receitas_operacionais	266	632,516,122	1,279,805,244	1,873,083	13,165,152,386
sobras	266	83,805,017	275,197,284	-708,226,808	1,856,914,876
pecld	266	-258,207,684	603,566,900	-6,021,173,359	-21,920

```
coop |>
dplyr::filter(big_four == 1) |>
dplyr::select(
  idade_em_2022,
  numero_agencias,
  total_de_cooperados,
  ativo_total,
  patrimonio_liquido,
  despesas_operacionais,
  receitas_operacionais,
  sobras,
  pecld
) %>%
stargazer::stargazer(
  ... = .,
  type = "text",
  title = "Estatística Descritiva da amostra de cooperativas de crédito singulares auditadas por Big Four",
  style = "aer",
  decimal.mark = ".",
  digits = 0
)
```

Estatística Descritiva da amostra de cooperativas de crédito singulares auditadas por Big Four

Statistic	N	Mean	St. Dev.	Min	Max
idade_em_2022	99	23	8	10	56
numero_agencias	99	13	19	1	118
total_de_cooperados	99	24,925	82,198	0	794,684
ativo_total	99	191,589,829,207	525,011,788,317	6,450,994,418	4,445,450,799,632
patrimonio_liquido	99	11,060,860,873	26,530,630,050	400,680,009	171,822,871,461
despesas_operacionais	99	-2,966,758,808	6,222,305,817	-42,527,929,368	-248,037,857

receitas_operacionais	99	3,687,521,815	7,751,230,023	292,141,748	53,567,939,894
sobras	99	690,040,006	2,029,670,201	-450,723,235	14,858,323,408
pecld	99	-1,775,911,501	4,688,592,034	-28,956,184,003	-56,045,651

---

Comparando das duas tabelas é possível observar que a quantidade de cooperativas de crédito auditada por *Big Four* é menor que as não auditadas, sendo 27% auditada e 70% não auditada por *Big Four*. Ainda, é possível observar que cooperativas auditadas por *Big Four* possuem idade média mais baixa, mas tamanho médio mais elevado em todas as variáveis.

Todas estas variáveis apresentaram diferenças estatisticamente significantes entre os grupos, ou seja, os dois grupos apresentam características distintas quanto a idade e tamanho.

## Metodologia

O primeiro passo foi corrigir o problema de tamanho das cooperativas notado pelos valores dos desvios da média, todas as variáveis financeiras (Ativo Total, Patrimônio Líquido, Despesas, Receitas, Sobras e PECLD) através da normalização por escalas em torno da média, com a fórmula:  $(\text{valor} - \min(\text{valor})) / (\max(\text{valor}) - \min(\text{valor}))$ . Não foram normalizadas por logaritmo por algumas variáveis possuírem valores negativos, que me fizeram optar por utilizar a mesma metodologia de normalização.

Este método também resolve problemas de distribuição das variáveis.

```
# Regressão logit

modelo.coop <- coop |>
  dplyr::mutate(
    escala_ativo_total = scale(ativo_total),
    escala_patrimonio_liquido = scale(patrimonio_liquido),
    escala_despesas_operacionais = scale(despesas_operacionais),
    escala_receitas_operacionais = scale(receitas_operacionais),
    escala_sobras = scale(sobras),
    escala_pecld = scale(pecld),
    filiacao = dplyr::case_when(filiacao != 0 ~ 1,
                               filiacao == 0 ~ 0)
  ) |>
  dplyr::select(
    big_four, idade_em_2022, numero_agencias, total_de_cooperados,
    escala_ativo_total, escala_patrimonio_liquido, escala_despesas_operacionais,
    escala_receitas_operacionais, escala_sobras, escala_pecld, filiacao
  )
```

Definida a mostra e os grupos a serem classificados, o modelo econométrico escolhido foi o Logit, que conforme Hair et al. (2005), é útil para situações nas quais se deseja prever a presença ou ausência de uma característica baseado em valores das variáveis independentes. Pode ser utilizada, por exemplo, para se mensurar a probabilidade do risco de crédito em situações de operação de vendas a prazo, empréstimos ou financiamentos. A probabilidade máxima pode ser estimada pela logit, após a transformação da variável dependente em variável de base logarítmica, permitindo que seja calculada a probabilidade de um certo evento acontecer.

Uma vantagem é a maior flexibilidade em seus pressupostos em relação a outras técnicas (como análise discriminante), como no caso de não precisar se preocupar com homogeneidade de variância e normalidade dos resíduos. Embora o resíduo também precise ter média zero, ausência de autocorrelação, correlação entre resíduos e variáveis explicativas e multicolinearidade.

```
log.audit <-
  modelo.coop |>
  glm(big_four ~ .,
      data = _,
      family = binomial
  )
```

Warning: glm.fit: probabilidades ajustadas numericamente 0 ou 1 ocorreu

Lembrando que, de acordo com Fávero e Belfiore (2017), a regressão logística binária estima, por máxima verossimilhança, não são os valores previstos da variável dependente, mas, sim, a probabilidade de ocorrência do evento em estudo para cada observação. No caso, a probabilidade de uma cooperativa de crédito ser auditada por uma *Big Four*.

A fórmula do modelo será:

$$BigFour = \alpha + \beta_1 Idade + \beta_2 NumeroAgencias + \beta_3 TotalCooperados + \beta_4 AtivoTotal + \beta_5 PatrimonioLiquido + \beta_6 Despesas + \beta_7 Receitas + \beta_8 Sobras + \beta_9 PECLD + \beta_{10} Filiacao$$

Após a estimação do modelo Logit, serão realizados os testes para verificar o nível de eficiência do modelo. Estes testes serão McFadden, para encontrarmos o chamado pseudo  $R^2$ , a matriz de confusão para verificarmos os níveis de acerto do modelo e a curva ROC para verificarmos o desempenho do modelo.

De acordo com Fávero e Belfiore (2017), como a variável dependente é qualitativa, não faz sentido discutirmos o percentual de sua variância que é explicado pelas variáveis preditoras, ou seja, em modelos de regressão logística não há um coeficiente de ajuste  $R^2$ , como nos modelos tradicionais de regressão estimados pelo método de mínimos quadrados ordinários. Então é realizado o teste do pseudo  $R^2$  proposto por McFadden e Domencich (1975) que indica uma relação entre a probabilidade e  $R$ , baseado na verossimilhança. Já a matriz de confusão nos mostra porcentagem de acertos do modelo, chamado de acurácia.

Por fim, será feita a análise da curva ROC, que é utilizada para categorizar uma variável numérica X em relação a uma variável categórica Y. A área abaixo da curva representa a probabilidade de que a curva ROC irá classificar a cooperativa corretamente. Ela varia de 0 a 1, sendo que 0.5 representa um modelo que seria completamente aleatório. Quanto mais próximo de 1, e consequentemente longe de 0.5, melhor o ajuste da curva ROC.

## Resultados

Os resultados da estimação do modelo Logit binomial pode ser observado abaixo:

```
# Descricao dos resultados
summary(log.audit)
```

Call:

```
glm(formula = big_four ~ ., family = binomial, data = modelo.coop)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.4692	-0.7399	-0.1133	0.2672	3.0954

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.109e+00	1.253e+00	-1.683	0.092293 .
idade_em_2022	-6.777e-02	2.605e-02	-2.602	0.009278 **

numero_agencias	-5.737e-02	2.773e-02	-2.069	0.038582	*
total_de_cooperados	1.213e-05	2.479e-05	0.489	0.624768	
escala_ativo_total	-1.979e+00	8.929e-01	-2.216	0.026707	*
escala_patrimonio_liquido	-1.162e+00	1.031e+00	-1.127	0.259817	
escala_despesas_operacionais	-4.524e+00	3.387e+00	-1.336	0.181635	
escala_receitas_operacionais	1.156e+00	3.304e+00	0.350	0.726383	
escala_sobras	2.116e+00	9.502e-01	2.227	0.025965	*
escala_pecld	1.938e+00	1.341e+00	1.445	0.148449	
filiacao	3.704e+00	1.010e+00	3.667	0.000246	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 426.67 on 364 degrees of freedom  
 Residual deviance: 254.38 on 354 degrees of freedom  
 AIC: 276.38

Number of Fisher Scoring iterations: 7

É possível observar que o modelo classificou a probabilidade da cooperativa de crédito optar por ser auditada por empresa comum. Com estes dados, a única variável capaz de explicar a adoção da auditoria por comum com intervalo de confiança de 99% é a filiação. A variável idade explica a adoção em 95% e as variáveis número de agência, ativo total e sobras (que são relativas ao tamanho) explicam o modelo a 90%. As demais variáveis não possuem efeito significativo na classificação.

A ordem de importância das variáveis para o resultado do modelo, independente se afeta positiva ou negativamente, é dada por:

```
# Importância de cada variável
caret::varImp(log.audit) |> dplyr::arrange(-Overall)
```

	Overall
filiacao	3.6665841
idade_em_2022	2.6016416
escala_sobras	2.2267415
escala_ativo_total	2.2157755
numero_agencias	2.0686202
escala_pecld	1.4450306
escala_despesas_operacionais	1.3357381
escala_patrimonio_liquido	1.1268239
total_de_cooperados	0.4891038
escala_receitas_operacionais	0.3499409

Em resumo: quando a empresa é filiada a uma cooperativa central e quanto mais velha, quanto maior for o ativo total, quanto mais agências esobras, maior as chances da cooperativa optar por ser auditada por uma empresa *Big Four*.

```
DescTools::PseudoR2(log.audit)
```

McFadden  
0.4038103

Além destes resultados, o pseudo  $R^2$  foi de 0,403, o que indicaria um poder explicativo muito bom para o modelo (McFadden e Domencich 1975) por estar entre 0,2 e 0,4.

Quando aplicamos a matriz de confusão para contar nos níveis de acertos, temos um bom resultado.

```
# Matriz de Confusão
glm.auditprob <- predict(log.audit, type = "response")
glm.auditprob <- ifelse(glm.auditprob > 0.5, "Big Four", "Comum")
audit <- ifelse(modelo.coop$big_four > 0.5, "Big Four", "Comum")
cm.audit <- table(glm.auditprob, audit)
caret::confusionMatrix(cm.audit)
```

Confusion Matrix and Statistics

```
      audit
glm.auditprob Big Four Comum
      Big Four      53      23
      Comum       46     243
```

```
      Accuracy : 0.811
      95% CI : (0.767, 0.8498)
No Information Rate : 0.7288
P-Value [Acc > NIR] : 0.0001637
```

```
      Kappa : 0.4842
```

```
McNemar's Test P-Value : 0.0080853
```

```
      Sensitivity : 0.5354
      Specificity : 0.9135
      Pos Pred Value : 0.6974
      Neg Pred Value : 0.8408
      Prevalence : 0.2712
      Detection Rate : 0.1452
      Detection Prevalence : 0.2082
      Balanced Accuracy : 0.7244
```

```
'Positive' Class : Big Four
```

A matriz de confusão apresentada indica que o modelo tem acurácia de 81% e um p-valor abaixo de 0,05, indicando robustez. Nesta matriz também é possível observar que o modelo previu corretamente 91% das cooperativas auditadas por empresa comum e apenas 53% das cooperativas auditadas por *Big Four*.

```
# Curva ROC
roc.audit <- pROC::roc(audit, glm.auditprob)
```

```
Setting levels: control = Big Four, case = Comum
```

```
Setting direction: controls > cases
```

```
roc.audit
```

Call:

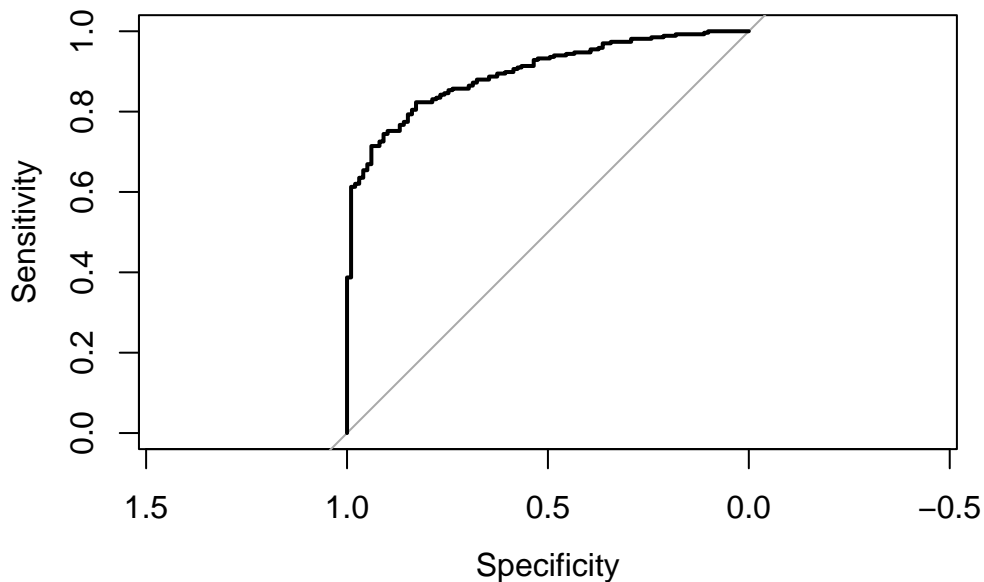
```
roc.default(response = audit, predictor = glm.auditprob)
```

```
Data: glm.auditprob in 99 controls (audit Big Four) > 266 cases (audit Comum).
```

```
Area under the curve: 0.896
```



```
pROC::plot.roc(roc.audit)
```



Quando plotamos a curva ROC temos que o modelo atingiu um nível satisfatório de classificação, sob 89%, bem próximo de 1. Com isso, podemos concluir que o modelo utilizado pode ser utilizado para prever as chances de uma cooperativa de crédito ser auditada por uma empresa *Big Four* ou não.

## Conclusão

Podemos concluir que apesar da importância de uma auditoria ser realizada por uma empresa com maior credibilidade no mercado, poucas cooperativas optam por contratar uma *Big Four*. Possivelmente, isso acontece devido a falta de mercado monitorando as cooperativas de crédito, uma vez que suas quotas não podem ser negociadas por terceiros.

Um fato interessante é que as chances de uma cooperativa ser auditada por *Big Four* aumentam quando ela está filiada a um sistema. Uma possível explicação para isso é que o sistema exige um nível maior de confiança nos relatórios contábeis para que a cooperativa possa fazer parte. Ainda, o estudo mostra que o tamanho e a idade podem justificar a escolha. Para estudos futuros, sugere-se utilizar outros métodos ou períodos maiores de análise, assim como inclusão de variáveis como tipo de cooperado ou região.

## Nota

Quando tentei normalizar as variáveis contábeis por  $\log()$ , tratando as variáveis negativas como  $\log(\text{var} - \min(\text{var}))$ , elas perderam o poder explicativo. Entretanto, o poder de explicação do modelo aumentou, fazendo com que apenas idade, quantidade de agências e filiação fossem significantes para classificar uma cooperativa como sendo auditada por *Big Four* ou não. Isso melhorou a capacidade de previsão do modelo de 89% para 91%.

## Referências

- Barton, D. 1983. «What is a cooperative?» *Cooperatives in agriculture* 7 (1).
- Fávero, Luiz Paulo, e Patrícia Belfiore. 2017. *Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®*. Elsevier Brasil.
- Hair, Joseph F, Rolph E Anderson, Ronald L Tatham, e William C Black. 2005. «Análise multivariada de dados. 5ª edição». Porto Alegre: Bookman.

- Jensen, Michael, e William Meckling. 1976. «Theory of the firm: Managerial behavior, agency costs and ownership structure». *Journal of financial economics* 3 (4): 305–60.
- Le, Thien, e Gerald J Lobo. 2020. «Audit Quality Inputs and Financial Statement Conformity to Benford's Law». *Journal of Accounting, Auditing & Finance*, 0148558X20930467.
- McFadden, Daniel L, e Tom Domencich. 1975. «Urban travel demand: A behavioral analysis».
- Schaefer, Vanessa, Sandro Augusto Martins Bittencourt, e Luana Zanetti Trindade Ferraz. 2022. «AUDITORIA INDEPENDENTE EM COOPERATIVAS AGROPECUÁRIAS: PERCEPÇÕES DOS RESPONSÁVEIS PELA GOVERNANÇA SOBRE MOTIVOS DA CONTRATAÇÃO». *Revista Gestão e Desenvolvimento* 19 (1): 30–56.
- Theodoro, Ricardo, Carlos Alberto Grespan Bonacim, e Davi Rogério de Moura Costa. 2021. «Identificação de Ações Discricionárias do Gestor em Cooperativas de Crédito: Uma aplicação da Lei de Benford». *Contabilidade Gestão e Governança* 24 (3): 331–48.