

# Aplicação do Índice de Hirschman-Herfindahl no setor financeiro brasileiro

Estudo para os anos de 2017 a 2021

Alexandre Barros

30 de agosto de 2022

## Introdução

O objetivo deste documento é apresentar, metodicamente e em etapas explicadas no R, a utilização de dados do setor financeiro brasileiro, disponibilizados pelo Banco Central do Brasil, nos [relatórios do IF.Período]]([https://www3.bcb.gov.br/ifPerÃnodo/#]).

Para tanto, além desta breve introdução, o documento se desdobrará, sempre explicando os comandos, no método e nos resultados aplicados com dados dos primeiros trimestres de 2017 a 2021, a fim de obter informações do grau de concentração neste setor.

## Base de Dados

Os dados do portal IF.Período, do BC, disponibilizam informações diversas, resumidas e desdobradas, de diversas instituições (cerca de 1500) e em diversos segmentos do setor financeiro. O primeiro passo corresponde a importação desses dados (obtidos em .csv) de maneira off-line, após-realizados o download.

No R, utilizamos as funções do pacote Dplyr de importação. Abaixo, usa-se a função *read.csv()* para importar do diretório os arquivos.

```
#2017 #
dados_17 <- read.csv("dados bancários 2017.csv", sep = ";", skip = 1) %>%
  select(1,6,17,28) %>% filter(X.5 != "") %>% mutate(Período = "03/2017")
#2018
dados_18 <- read.csv("dados bancários 2018.csv", sep = ";", skip = 1) %>%
  select(1,6,17,28) %>% filter(X.5 != "") %>% mutate(Período = "03/2018")
#2019
dados_19 <- read.csv("dados bancários 2019.csv", sep = ";", skip = 1) %>%
  select(1,6,17,28) %>% filter(X.5 != "") %>% mutate(Período = "03/2019")
#2020
dados_20 <- read.csv("dados bancários 2020.csv", sep = ";", skip = 1) %>%
  select(1,6,17,28) %>% filter(X.5 != "") %>% mutate(Período = "03/2020")
#2021
dados_21 <- read.csv("dados bancários 2021.csv", sep = ";", skip = 1) %>%
  select(1,6,17,28) %>% filter(X.5 != "") %>% mutate(Período = "03/2021")
```

Além da importação, selecionou-se apenas as colunas ou variáveis: 1) nome da instituição, 2) subsetor da instituição, 3) operações de crédito líquidas, e 4) Ativo total das instituições; isto para cada período.

Os comandos abaixo organizam os dados em uma base consolidada para sua utilização.

1 - Consolidando todos os dados em um só data.frame.

```
Dados.com <- bind_rows(dados_17, dados_18, dados_19, dados_20, dados_21)
```

2 - Alterando o nome das variáveis.

```
names(Dados.com) <- c("Instituições",  
                      "Tipo",  
                      "Operações de Crédito",  
                      "Ativo total",  
                      "Período")
```

3 - Retirando demais objetos do R (retirando excessos).

```
gdata::keep(Dados.com, sure = T)
```

## Analizando os dados

Agora que consolidamos os dados, para os respectivos períodos, no objeto a que chamamos *Dados.com*, podemos analisar preliminarmente estes dados.

O próximo passo é será revisar os dados. Ou seja, ocorre que no processo de importação, o R nem sempre reconhece o tipo correto de dados. É dever nosso completar esses dados. Abaixo o script responsável por aprimorar as variáveis existentes, usando a função *mutate()*, que cria novas variáveis a partir de variáveis existentes.

Como se pode ver, adicionamos os Subsetores de cada setor de maneira a melhor analisar os dados, com o auxílio da função *case\_when*, que usa de operações lógicas para identificar corretamente o subsetor da instituição.

```
Dados.com %<>% mutate( #  
  `Ativo total` = as.numeric(str_replace_all(`Ativo total`, "\\.", ""))/1000,  
  
  `Operações de Crédito` = as.numeric(str_replace_all(`Operações de Crédito`, "\\.", ""))/1000,  
  
  Subsetor = case_when(  
    Tipo == "b1" ~ "Bancos comerciais",  
    Tipo == "b2" ~ "Bancos de investimento\n e câmbio",  
    Tipo == "b3C" |  
    Tipo == "b3S" ~ "Cooperativas Singulares\n e Confederações",  
    Tipo == "b4" ~ "Bancos\n de desenvolvimento",  
    TRUE ~ "Não bancários\n de crédito e capital"  
  ))
```

Ainda no pacote Dplyr, resumizamos as informações conforme o script abaixo e mostramos seu resultado em uma tabela do tipo Kable.

Cada uma das variáveis criadas contém os valores e informações analisados, neste caso, o total de crédito movimentado pelas instituições e seu ativo acumulado. Podemos mostrar esses dados, como mencionado, usando um kable, como se vê abaixo.

Subsetor	Crédito			Ativo			Total das operações de crédito
	Número de instituições	Média de ativos	Desvio Padrão	Total dos ativos das instituições	Média de ativos i	Desvio Padrão i	
Bancos comerciais	121	58930.5293	227646.2729	7130594.0	17154.48608	81475.1343	2075692.82
Bancos comerciais	121	58444.0807	223699.4890	7071733.8	16857.67724	79820.3831	2039778.95
Bancos comerciais	120	62584.5057	234742.8157	7510140.7	17480.89597	80042.5492	2097707.52
Bancos comerciais	118	71688.6397	258908.6862	8459259.5	19228.93591	84359.0386	2269014.44
Bancos comerciais	122	76856.2564	276577.0460	9376463.3	22371.07794	95243.7314	2729271.51
Bancos de investimento e câmbio	49	6099.7934	14953.8691	298889.9	2057.51961	3684.9182	100818.46
Bancos de investimento e câmbio	47	6287.2402	14600.7006	295500.3	2259.36345	3901.2679	106190.08
Bancos de investimento e câmbio	49	6409.4601	14316.2413	314063.5	2513.73073	4498.6350	123172.81
Bancos de investimento e câmbio	50	7609.7054	16950.2313	380485.3	2974.40478	5382.8850	148720.24
Bancos de investimento e câmbio	50	7501.1364	18408.2504	375056.8	2872.49282	5689.2304	143624.64
Bancos de desenvolvimento	4	222844.3417	428854.9195	891377.4	84648.26225	156240.4318	338593.05
Bancos de desenvolvimento	4	220123.2657	423484.2391	880493.1	76407.34975	139933.8657	305629.40
Bancos de desenvolvimento	4	218613.4378	420671.5729	874453.8	72797.95600	133418.3690	291191.82
Bancos de desenvolvimento	4	191153.0200	366384.5756	764612.1	66662.08700	121633.8340	266648.35
Bancos de desenvolvimento	4	200127.8682	382916.9837	800511.5	68933.38025	125228.9071	275733.52
Cooperativas Singulares e Confederações	1041	224.2726	653.5361	232570.7	76.49042	172.6136	79320.56
Cooperativas Singulares e Confederações	1002	263.3706	722.4167	263634.0	92.57099	200.4928	92663.56
Cooperativas Singulares e Confederações	946	318.4973	788.8557	300661.5	123.08610	257.8310	116193.28
Cooperativas Singulares e Confederações	900	378.1611	828.6549	339588.7	162.43990	331.2519	145871.03
Cooperativas Singulares e Confederações	875	525.4639	1103.6114	457153.6	233.88032	468.3193	203475.88
Não bancários de crédito e capital	384	1477.2030	10223.2066	561337.2	224.20178	2011.9586	85196.68
Não bancários de crédito e capital	375	1494.4653	7523.9598	557435.5	253.56360	2341.0042	94579.22
Não bancários de crédito e capital	375	1538.1929	7311.8353	570669.6	284.14846	2671.1375	105419.08
Não bancários de crédito e capital	384	1547.3302	7122.8476	589532.8	310.38381	2945.1893	118256.23
Não bancários de crédito e capital	414	1576.9031	7262.1632	635491.9	209.05560	2251.0614	84249.41

```
knitr::kable(select(Dados.sum, -2), format="latex", booktabs = TRUE) %>%
  kable_styling(latex_options="scale_down") %>%
  add_header_above(c(" ", "Crédito" = 3, "Ativo" = 3)) %>%
  row_spec(row = 5, hline_after = T) %>%
  row_spec(row = 10, hline_after = T) %>%
  row_spec(row = 15, hline_after = T) %>%
  row_spec(row = 20, hline_after = T)
```

## Metodologia

### Relação de Concentração

A Relação de Concentração (CR) mede a parcela de participação de um número determinado de firmas no contexto geral da indústria examinada. A fórmula abaixo ilustra a aplicação da técnica numa situação em que se leva em conta as quatro maiores empresas do setor:

$$CR4 = \sum_{i=4}^k S_i$$

Aqui, se aplicará a Relação de Concentração para as quatro e oito maiores empresas de cada segmento. A tabela abaixo ajuda a compreender os resultados.

Níveis de Mercado CR	CR4	CR8
Altamente Concentrado	i > 75%	i > 90%
Alta Concentração	65% < i < 75%	85% < i < 90%
Concentração moderada	50% < i < 65%	70% < i < 85%
Baixa Concentração	35% < i < 50%	45% < i < 70%
Ausência de concentração	i < 35%	i < 45%
Claramente Atomístico	i = 2%	

## Índice de Hirshman-Herfindahl (HHI)

Outra técnica para se medir concentração de mercado é o Índice de Herfindahl-Hirschman (HHI), em que são somados os quadrados da parcela de participação de cada empresa, levando-se em conta todas as empresas conforme ilustra a expressão abaixo:

$$HHI = \sum_{i=1}^k S_i^2$$

Ao contrário da relação CR, o Índice HHI leva em conta o tamanho relativo das empresas ao elevar a parcela de participação de cada uma delas ao quadrado. A tabela abaixo ajuda a compreendê-lo:

Índice	Classificação do mercado
HHI < 1000	Não concentrado
HHI > 1000 porém < 1800	Moderadamente concentrado
HHI > 1800	Altamente concentrado

No ambiente do R, vamos utilizar de um mecanismo provido pela ferramentas em que fixamos um grupo (isto é, um eixo) e obtemos apenas os maiores valores selecionados, nesta caso, as 4 e 8 maiores instituições, respectivamente, que atuam em determinado segmento do setor financeiro.

Segue-se, sequencialmente, os scripts que realizam esta tarefa tanto para as operações de crédito quanto para os ativos detidos pelos bancos. Os comandos abaixo têm, em essência, as mesmas funções, sendo estas as responsáveis por determinada tarefa nos dados.

- *group\_by()*, responsável por fixar os eixos ou grupos analisados;
- *slice\_max()*, obtém os maiores valores de cada grupo, recebendo como argumento a variável e quantidade de posições;
- *merge()*, mescla bases diferentes com características (ou variáveis) comuns, similar a uma chave primária;
- *mutate()*, cria novas variáveis a partir da base selecionada (cf. comando pipe);
- *summarise()*, agrega os valores conforme o grupo e operação selecionada, usamos *sum()* para somar os valores das maiores instituições.

1 - Obtendo a participação das 4 maiores empresas por setor, por período, para o total de ativos financeiros.

```
CR4_a <- Dados.com %>%
  group_by(Subsetor, Período) %>%
  slice_max(`Ativo total`, n = 4) %>%

  merge(Dados.sum, by = c("Período", "Subsetor"), all.x = T) %>%
  mutate(Participação = (`Ativo total`/`Total dos ativos das instituições`)*100) %>%
  group_by(Subsetor, Período) %>%
  summarise(`RC Ativo` = sum(Participação))
```

2 - Obtendo a participação das 4 maiores empresas por operações de créditos realizadas no período.

```
CR4_c <- Dados.com %>%
  group_by(Subsetor, Período) %>%
  slice_max(`Operações de Crédito`, n = 4) %>%
```

```
merge(Dados.sum, by = c("Período", "Subsetor"), all.x = T) %>%
mutate(Participação = (`Operações de Crédito` /
                      `Total das operações de crédito`) * 100) %>%
group_by(Subsetor, Período) %>%
summarise(`RC Crédito` = sum(Participação)) %>% mutate(Indicador = "CR4")

# juntando as duas bases
CR4 <- merge(CR4_a, CR4_c, by = c("Período", "Subsetor"))
```

3 - Obtendo a participação das 8 maiores empresas por setor (CR8), por período, para o total de ativos financeiros.

```
CR8_a <- Dados.com %>%
group_by(Subsetor, Período) %>%
slice_max(`Ativo total`, n = 8) %>%

merge(Dados.sum, by = c("Período", "Subsetor"), all.x = T) %>%
mutate(Participação = (`Ativo total` /
                      `Total dos ativos das instituições`) * 100) %>%
group_by(Subsetor, Período) %>%
summarise(`RC Ativo` = sum(Participação))
```

4 - Obtendo a participação das 4 maiores empresas por operações de créditos realizadas no período.

```
CR8_c <- Dados.com %>%
group_by(Subsetor, Período) %>%
slice_max(`Operações de Crédito`, n = 8) %>%

merge(Dados.sum, by = c("Período", "Subsetor"), all.x = T) %>%
mutate(Participação = (`Operações de Crédito` /
                      `Total das operações de crédito`) * 100) %>%
group_by(Subsetor, Período) %>%
summarise(`RC Crédito` = sum(Participação)) %>% mutate(Indicador = "CR8")

CR8 <- merge(CR8_a, CR8_c, by = c("Período", "Subsetor"))
```

5 - Consolidando todas as informações em um só data.frame.

```
OC <- bind_rows(CR4, CR8)

OC %>% pivot_longer(cols = `RC Ativo`:`RC Crédito`,
                    values_to = "Participação",
                    names_to = "CR")

# retirando excessos
gdata::keep(Dados.com, OC, Dados.sum, sure = T)
```

Pronto! Consolidamos a base correspondente a relação de concentração. Agora é tarefa nossa realizar as operações para obter o Índice Hirschman-Herfindahl.

Como mencionado antes, este índice toma como base a participação das empresas e as eleva ao quadrado, somando todos os valores da base. Para os dados que dispomos, esta operação é similar àquela realizada

para a relação de concentração, com a diferença da operação matemática e do uso de todas as instituições dos segmentos, não apenas os maiores.

Desta forma, realizamos isso em três etapas simples:

1 - Sumarizando os valores para cada empresa por segmento (participação individual)<sup>1</sup>.

```
HHI <- Dados.com %>%
  group_by(Período, Subsetor) %>%
  summarise(HHI_crédito = (`Operações de Crédito` /
    sum(`Operações de Crédito`, na.rm = T)*100)^2,
    HHI_ativo = (`Ativo total` /
    sum(`Ativo total`, na.rm = T)*100)^2)
```

2 - Agregando os valores por segmento, obtendo o HHI pela função *sum()* (somatório).

```
HHI %<>%
  group_by(Período, Subsetor) %>%
  summarise(Crédito = sum(HHI_crédito, na.rm = T),
    `Ativo` = sum(HHI_ativo, na.rm = T))
```

3 - Por último, nós juntamos as variáveis para facilitar a visualização no gráfico.

## Resultados

Para terminar esta

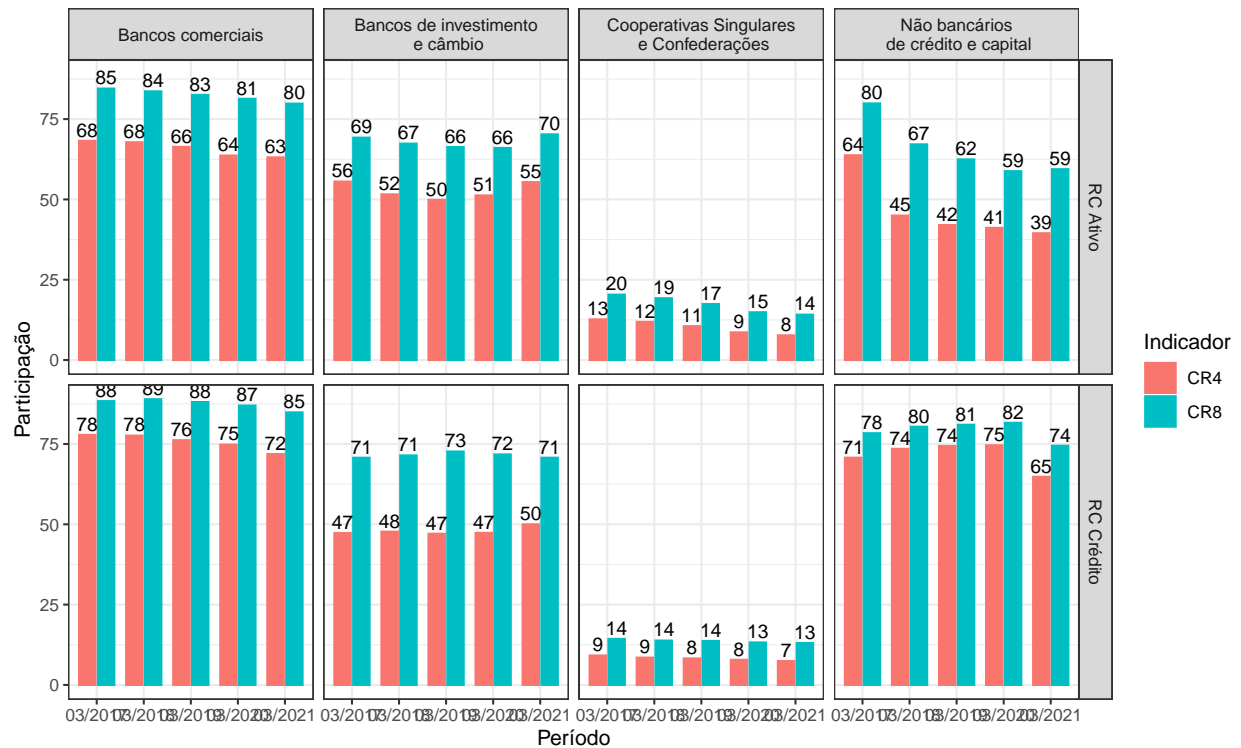
```
p <-
  filter(OC, Subsetor != "Bancos\n de desenvolvimento") %>%
  ggplot(aes(x = Período, y = Participação)) +
  geom_bar(
    aes(color = Indicador, fill = Indicador),
    stat = "identity", position = position_dodge(0.8),
    width = 0.7
  ) +
  theme(axis.text.x = element_text(size = 10,
    color = 'black',
    angle = 90,
    vjust = .3))

p <- p + geom_text(
  aes(label = round(Participação,0), group = Indicador),
  position = position_dodge(0.8),
  vjust = -0.3, size = 3.5
)

p + facet_grid(CR~Subsetor) + theme_bw()
```

---

<sup>1</sup>Perceba que aqui nós já aplicamos a exponenciação da participação



```
test <-
ggplot(HHI, stat = 'identity', position = "dodge",
  aes(x = Período, y = Índice, group = Tipo, label = round(Índice,0), fill = Tipo)) +
  geom_line(alpha = 0.8, stroke = 0, size = 1.1) +
  geom_point() +
  geom_label(segment.colour="black", size=3.1) +
  facet_wrap(~Subsetor, scales = "free_y") +
labs(title = "Concentração Bancária - Índice de Hirschman-Herfindahl (HHI)",
  subtitle = "1º tri de 2017-2021",
  x = "Período",
  y = "Pontos HHI") +
theme_bw() +
  theme(axis.title = element_text(size = 12),
    axis.title.y = element_text(size = 11.5),
    axis.text.x = element_text(size = 10, color = 'black', angle = 90, vjust = .3),
    axis.text.y = element_text(size = 11, color = 'black', face = 'bold'),
    legend.title = element_text(size = 12),
    legend.position= 'bottom' ,
    legend.text=element_text(size=13),
    legend.background = element_rect(fill="ghostwhite", size=1),
    axis.ticks = element_line(colour = "grey70", size = 0.2),
    panel.grid.major = element_line(colour = "grey70"),
    strip.text = element_text(face = 'bold'))

test
```

Concentração Bancária – Índice de Hirschman–Herfindahl (HHI)  
1º tri de 2017–2021

