

Análise de Conjuntura usando o R

Núcleos de Inflação

Vítor Wilher

Cientista de Dados



Plano de Voo

Introdução

Conceitos

Núcleos de inflação

O pacote {nucleos}

Curso de Análise de Conjuntura usando o R - Versão 5.0

Introdução

Nesta aula aprenderemos a **constuir as medidas de núcleos de inflação** comumente acompanhadas pelo Banco Central do Brasil, quais sejam:

- **Núcleos de exclusão:**
 - IPCA-EX0
 - IPCA-EX1
 - IPCA-EX2
 - IPCA-EX3

- **Núcleos estatísticos:**
 - IPCA-MA
 - IPCA-MS
 - IPCA-DP

Conceitos

O **Índice de Preços ao Consumidor Amplo (IPCA)**¹ é aquele utilizado pelo Banco Central do Brasil (BCB) para cumprir o determinado no decreto 3.088, de 21 de junho de 1999, que estipulou o *regime de metas para inflação* no país.

O IPCA é construído de forma hierarquizada, sendo dividido em **grupos, subgrupos, itens e subitens**. Atualmente são nove os grupos: alimentação e bebidas, habitação, artigos de residência, vestuário, transportes, saúde e cuidados pessoais, despesas pessoais, educação e comunicação.

¹Consulte informações sobre o IPCA em:

<https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html>

Conceitos

Todos os meses o IBGE calculará a variação do preço de cada um dos **subitens** em relação ao mês anterior, bem como verificará o seu peso no índice cheio.

O peso de cada um desses grupos no índice cheio é definido mensalmente, de acordo com os gastos das famílias em diferentes bens e serviços. Estes gastos, por sua vez, são captados pela **Pesquisa de Orçamento Familiar (POF)**, também do IBGE.²

O somatório do produto entre variação mensal e peso de todos os subitens dará a **variação mensal do índice cheio**.

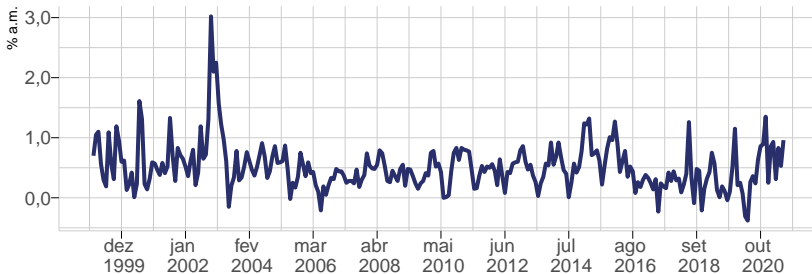
²Para maiores informações, consulte:

<https://www.ibge.gov.br/estatisticas/sociais/saude/24786-pesquisa-de-orcamentos-familiares-2.html>

Conceitos

Com o gráfico abaixo, podemos avaliar o que chamamos de **inflação cheia**, representada pelo comportamento de todos os subitens.

Inflação medida pelo IPCA



Fonte: analisemacro.com.br com dados do IBGE.

Classificações do IPCA

Para algumas análises, pode ser interessante **reunir os subitens e itens do IPCA em determinadas classificações**, que não aquelas representadas pelos nove grupos divulgados pelo IBGE. Por exemplo, dentro do índice existem certos **preços que são regulados**, isto é, seus reajustes precisam ser autorizados pelo governo. Esses preços, chamados de **monitorados ou administrados**, seguem uma lógica diferente de **preços de mercado**, determinados pelo confronto entre oferta e demanda.

Assim, talvez seja melhor isolar esses preços e modelá-los conforme sua dinâmica própria, do que construir uma modelagem em cima do índice cheio.³

³Como veremos ao longo do nosso curso, de fato, esse é o modo que o Banco Central operacionaliza os seus modelos.

Classificações do IPCA

A tabela abaixo⁴ ilustra as diversas **classificações construídas pelo próprio Banco Central**, assim como suas mudanças nos pesos entre as duas últimas POF.

Tabela 6 – Classificações do IPCA: novos pesos

	Pesos em janeiro de 2018		%
	POF 2008-2009	POF 2017-2018	
	Classificação atual	Nova classificação	
IPCA	100,00	100,00	
Livres	74,72	74,91	
Alimentos	15,67	13,11	
Serviços	35,82	37,22	
Subjacente	22,44	21,66	
Ex-Subjacente	13,38	15,55	
Bens industriais	23,24	24,59	
Duráveis	7,83	10,82	
Semi-duráveis	7,74	6,29	
Não-duráveis	7,67	7,48	
Monitorados	25,28	25,09	
Livres	74,72	74,91	
Comercializáveis	33,93	30,87	
Não-comercializáveis	40,80	44,04	
Monitorados	25,28	25,09	

⁴Retirada do Estudo Especial nº 69/2019 do BCB:

Classificações do IPCA

Dessa forma, vale pontuar que os **fatores que influenciam a inflação de serviços não serão os mesmos dos que influenciam a inflação de bens comercializáveis**, por exemplo, ainda que ambas classificações respondam a aspectos comuns.

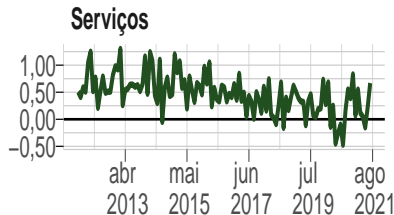
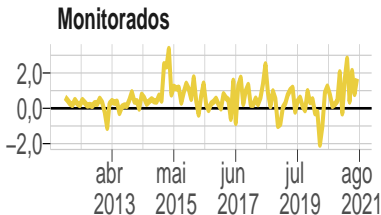
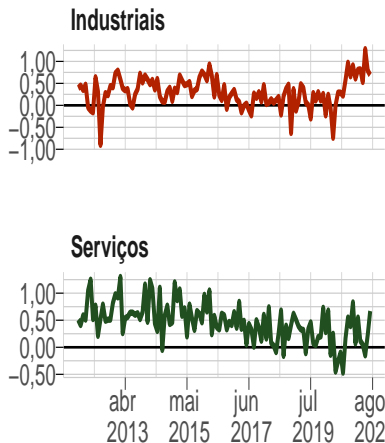
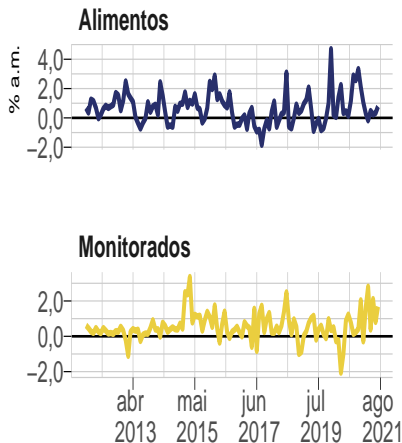
De fato, o estado da arte em modelagem e previsão leva em consideração a inflação desagregada e não a inflação cheia justamente por esse motivo.

Utilizaremos estas classificações para **construir** as medidas de núcleos da inflação.

Classificações do IPCA

Abaixo o comportamento de algumas das classificações do BCB:

Classificações do IPCA



Núcleos de inflação

Como pode ser imaginado, e confirmado, alguns itens do IPCA podem apresentar **variações mais infrequentes ou voláteis**, de modo que o Banco Central procura capturar e acompanhar a tendência inflacionária ao longo do tempo criando medidas auxiliares: **os núcleos do IPCA**.

Nesta aula focaremos na **análise e construção das sete medidas de núcleos** destacadas no início e comumente acompanhadas nos últimos anos pelo BCB e analistas de inflação⁵. Entretanto, pontua-se que desde julho de 2020 o BCB acompanha apenas cinco medidas de núcleos: EX0, EX3, MS, DP e P55.

⁵Deve-se observar que o conjunto de medidas divulgadas pelo BCB tende a sofrer alterações ao longo do tempo tanto pela inclusão de novas medidas como pela revisão de medidas que eram anteriormente divulgadas.

Núcleos de inflação

Descrição das medidas de núcleos divulgadas pelo BCB
(classificação a partir de janeiro/2020⁶):

Tipo	Núcleo	Descrição
Exclusão	IPCA-EX0	Exclui Alimentação no domicílio e Monitorados.
Exclusão	IPCA-EX1	Exclui Cereais, leguminosas e oleaginosas; Tubérculos, raízes e legumes; Açúcares e derivados; Hortaliças e Verduras; Frutas; Carnes; Pescados; Aves e ovos; Leite e derivados; Óleos e gorduras; Combustíveis (domésticos); Combustíveis (veículos).
Exclusão	IPCA-EX2	Exclui Cereais, leguminosas e oleaginosas; Farinhas, féculas e massas; Tubérculos, raízes e legumes; Açúcares e derivados; Hortaliças e Verduras; Frutas; Carnes; Pescados; Aves e ovos; Leite e derivados; Óleos e gorduras; Sal e condimentos; Aparelhos eletroeletrônicos; Automóvel novo; Automóvel usado; Etanol; Fumo; Serviços Ex-Subjacente; Monitorados.
Exclusão	IPCA-EX3	Exclui Alimentação no domicílio; Aparelhos eletroeletrônicos; Automóvel novo; Automóvel usado; Etanol; Fumo; Serviços Ex-subjacente; Monitorados.
Média aparada	IPCA-MA	Simétrica (20% em cada cauda) sem itens suavizados.
Média aparada	IPCA-MS	Simétrica (20% em cada cauda) com itens suavizados: Combustíveis domésticos, Energia elétrica residencial, Transporte público, Combustíveis (veículos), Serviços pessoais, Fumo, Cursos regulares, Cursos diversos e Comunicação.
Dupla ponderação	IPCA-DP	Pesos baseados na volatilidade relativa (48 meses) e nos pesos originais do IPCA, reduzindo a importância dos componentes mais voláteis.

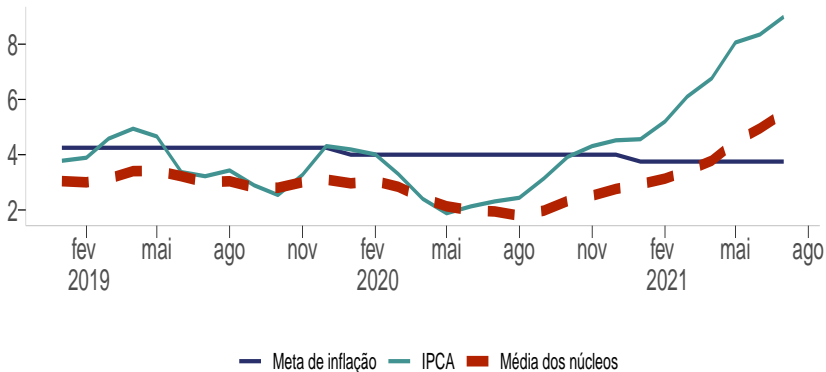
⁶Retirada do Estudo Especial nº 69/2019 do BCB:

Núcleos de inflação

Historicamente o comportamento dos núcleos tende a ser menos oscilante em relação ao IPCA:

Evolução dos núcleos vs. IPCA

Varição % acumulada em 12 meses



Núcleos de inflação

Conforme visto, os núcleos de exclusão são os mais simples metodologicamente. Já os estatísticos requerem alguns procedimentos e detalhes importantes.

Vamos esclarecer esses pontos, explorando em detalhes a **metodologia por trás de cada núcleo**, antes de ir para a prática.

Núcleos de inflação

O **núcleo por médias aparadas (IPCA-MA)** exclui os itens cuja variação mensal se situe, na distribuição, acima do percentil 80 ou abaixo do percentil 20. Os 60% restantes são utilizados para calcular a variação mensal do núcleo.

O procedimento de cálculo mensal consiste em:

- 1) ordenar os itens em ordem crescente de variação: $x_1 < x_2 < \dots < x_N$;
- 2) calcular a série acumulada de pesos $a(k)$, k de 1 até N , de acordo com o novo ordenamento;
- 3) cortar os itens que satisfaçam $a(k) < 0,20$ e $a(k) > 0,80$;
- 4) reponderar os itens remanescentes e a partir de suas variações calcular o núcleo.

Núcleos de inflação

Quando $a(k) < 0,20$ e $a(k+1) > 0,20$, o item $k+1$ entra no cálculo com peso equivalente à diferença do peso acumulada e 0,20. Por exemplo, se o peso acumulado atingir 0,22 no item $k+1$, o seu peso na reponderação passa a ser 0,02. O mesmo se aplica ao item k quando $a(k) < 0,80$ e $a(k+1) > 0,80$.

Obs: Os valores de cada período consideram a classificação e os pesos oficiais vigentes à época.

Núcleos de inflação

O **núcleo de médias aparadas com suavização (IPCA-MS)** exclui 20% da variação de preços dos itens do IPCA de ambas as caudas da distribuição mensal, após a suavização de componentes com variações infrequentes e de grandes magnitudes.

O procedimento de cálculo mensal consiste em:

- 1) suavizar os itens específicos*, calculando para eles a variação dos últimos 12 meses
- 2) ordenar estes itens recalculados e os demais em ordem crescente de variação $x_1 < x_2 < \dots < x_N$;
- 3) calcular a série acumulada de pesos $a(k)$, k de 1 até N , de acordo com o novo ordenamento;
- 4) cortar os itens que satisfaçam $a(k) < 0,20$ e $a(k) > 0,80$;
- 5) reponderar os itens remanescentes e a partir de suas variações calcular o núcleo.

Núcleos de inflação

Quando $a(k) < 0,20$ e $a(k+1) > 0,20$, o item $k+1$ entra no cálculo com peso equivalente à diferença do peso acumulada e 0,20. Por exemplo, se o peso acumulado atingir 0,22 no item $k+1$, o seu peso na reponderação passa a ser 0,02. O mesmo se aplica ao item k quando $a(k) < 0,80$ e $a(k+1) > 0,80$.

*Itens suavizados: Combustíveis domésticos, Energia elétrica residencial, Transporte público, Combustíveis (veículos), Serviços pessoais, Fumo, Cursos regulares, Cursos diversos e Comunicação.

Obs: Os valores de cada período consideram a classificação e os pesos oficiais vigentes à época.

Núcleos de inflação

O **núcleo de dupla ponderação (IPCA-DP)** ajusta os pesos originais de cada item de acordo com sua volatilidade relativa, reduzindo a importância de componentes mais voláteis.

O procedimento de cálculo mensal consiste em:

- 1) calcular para cada item o desvio padrão da série de diferenças entre a sua variação e a do IPCA nos últimos 48 meses $\sigma(k)$;
- 2) calcular novos $p(k)^* = [p(k)/\sigma(k)]/[Sum(i) p(i)/\sigma(i)]$;
- 3) utilizar os pesos p^* para ponderar as variações de preço dos diferentes itens do IPCA.

Obs: Os valores de cada período consideram a classificação e os pesos oficiais vigentes à época.

O pacote {nucleos}

Contextualizado a importância e conceituação sobre o tema, nos voltaremos agora para o aspecto prático de **como construir os núcleos de inflação usando o R**.

O procedimento não é trivial, mas em resumo:

- Criamos um projeto com um pacote de R localmente;
- No pacote criamos a estrutura de funções e objetos úteis;
- Disponibilizamos o pacote no GitHub para que possa ser instalado/utilizado.

Focaremos aqui em entender **o que as funções do pacote fazem**, ou seja, como o código escrito gera os núcleos de inflação similar ao divulgado pelo BCB⁷.

⁷Confira as séries calculadas pelo BCB no SGS:

<https://www3.bcb.gov.br/sgspub/>

O pacote {nucleos}

No console do RStudio, instale e carregue o pacote com:

```
# Instalar
if(!require("remotes")) install.packages("remotes")
remotes::install_github(
  repo      = "analisemacropro/nucleos",      # repositório privado
  auth_token = "digite aqui a chave de token"
)

# Carregar
library(nucleos)
```

O pacote {nucleos}

Em resumo, o pacote possui as seguintes **funcionalidades**⁸:

Tipo	Núcleo	Função
Exclusão	IPCA-EX0	nucleos::core_adhoc()
Exclusão	IPCA-EX1	nucleos::core_adhoc()
Exclusão	IPCA-EX2	nucleos::core_adhoc()
Exclusão	IPCA-EX3	nucleos::core_adhoc()
Média aparada	IPCA-MA	nucleos::core_ma()
Média aparada	IPCA-MS	nucleos::core_ms()
Dupla ponderação	IPCA-DP	nucleos::core_dp()

⁸Consulte em detalhes a documentação do pacote em:
<https://analisemacropro.github.io/nucleos/>

O pacote {nucleos}

Uma vez instalado o pacote, vamos demonstrar como utilizá-lo. Primeiro vamos **coletar os dados desagregados do IPCA**, ou seja, os subitens, itens, subgrupos e grupos do índice:

```
# Carregar pacotes adicionais para exemplos
if(!require("pacman")) install.packages("pacman")
pacman::p_load("dplyr", "magrittr")

# Coletar e tratar dados desagregados do IPCA
ipca_core <- nucleos::get_ipca(table = c(1419, 7060), period = "all") %>%
  nucleos::group_desc()

# Resultado
ipca_core

## # A tibble: 53,227 x 8
##   date      table code desc                pct_change weight snipc group
##   <date>    <chr> <dbl> <chr>                <dbl>    <dbl> <chr> <chr>
## 1 2012-01-01 1419 7169 Índice geral                0.56 100    ""    Geral
## 2 2012-01-01 1419 7170 1.Alimentação e beb-          0.86 23.1  "1"    Grupo
## 3 2012-01-01 1419 7171 11.Alimentação no d-         0.68 15.2  "11"   Subgr-
## 4 2012-01-01 1419 7172 1101.Cereais, legum-         5.44 0.822 "1101" Item
## 5 2012-01-01 1419 7173 1101002.Arroz                1.32 0.505 "1101~ Subit-
## 6 2012-01-01 1419 7175 1101051.Feijão - mu-         16.0 0.0216 "1101~ Subit-
## 7 2012-01-01 1419 7176 1101052.Feijão - pr-         5.42 0.0537 "1101~ Subit-
## 8 2012-01-01 1419 7177 1101053.Feijão - ma-         1.17 0.0336 "1101~ Subit-
## 9 2012-01-01 1419 12222 1101073.Feijão - ca-        15.1 0.207 "1101~ Subit-
## 10 2012-01-01 1419 41128 1101075.Feijão - br-        NA    NA    "1101~ Subit-
## # ... with 53,217 more rows
```

O pacote {nucleos}

Com os dados coletados, podemos facilmente calcular qualquer um dos 7 núcleos. Por exemplo, no caso do IPCA-EX0 (e de todos os núcleos de exclusão) usamos a função `core_adhoc()`:

```
# Calcular núcleo IPCA-EX0
ipca_ex0 <- nucleos::core_adhoc(
  core   = "ex0",           # pode ser também ex1, ex2, ex3
  data   = ipca_core,
  change = "pct_change",  # nome da coluna com variações mensais
  weight = "weight",      # nome da coluna dos pesos
  code   = "code",        # nome da coluna de identificação do código do elemento
  date   = "date"         # nome da coluna de datas
)
```

Resultado

```
ipca_ex0

## # A tibble: 115 x 2
##   date       ipca_ex0
##   <date>     <dbl>
## 1 2012-01-01    0.57
## 2 2012-02-01    0.65
## 3 2012-03-01    0.21
## 4 2012-04-01    0.71
## 5 2012-05-01    0.26
## 6 2012-06-01   -0.1
## 7 2012-07-01    0.44
## 8 2012-08-01    0.39
## 9 2012-09-01    0.39
## 10 2012-10-01    0.49
## # ... with 105 more rows
```


O pacote {nucleos}

Para os núcleos estatísticos existe uma função dedicada ao cálculo de cada um. No caso do núcleo de médias aparadas sem suavização basta rodar:

```
# Núcleo IPCA-MA
ipca_ma <- ipca_core %>%
  dplyr::filter(group == "Item") %>%
  nucleos::core_ma(., "pct_change", "weight", "date")
```

```
# Resultado
```

```
ipca_ma
```

```
## # A tibble: 115 x 2
##   date      ipca_ma
##   <date>    <dbl>
## 1 2012-01-01  0.36
## 2 2012-02-01  0.24
## 3 2012-03-01  0.27
## 4 2012-04-01  0.5
## 5 2012-05-01  0.31
## 6 2012-06-01  0.3
## 7 2012-07-01  0.35
## 8 2012-08-01  0.41
## 9 2012-09-01  0.46
## 10 2012-10-01  0.4
## # ... with 105 more rows
```

O pacote {nucleos}

Núcleo de médias aparadas com suavização:

```
# Núcleo IPCA-MS
ipca_ms <- ipca_core %>%
  dplyr::filter(group == "Item") %>%
  nucleos::core_ms(., "pct_change", "weight", "code", "date")
```

Resultado

ipca_ms

```
## # A tibble: 104 x 2
##   date      core_ms
##   <date>    <dbl>
## 1 2012-12-01  0.56
## 2 2013-01-01  0.69
## 3 2013-02-01  0.46
## 4 2013-03-01  0.47
## 5 2013-04-01  0.53
## 6 2013-05-01  0.49
## 7 2013-06-01  0.36
## 8 2013-07-01  0.28
## 9 2013-08-01  0.41
## 10 2013-09-01  0.42
## # ... with 94 more rows
```

O pacote {nucleos}

Núcleo de dupla ponderação:

```
# Núcleo IPCA-DP
ipca_dp <- ipca_core %>%
  nucleos::core_dp(., "pct_change", "weight", "code", "group", "date")

# Resultado
ipca_dp

## # A tibble: 67 x 2
##   date      core_dp
##   <date>    <dbl>
## 1 2016-01-01  0.87
## 2 2016-02-01  0.78
## 3 2016-03-01  0.49
## 4 2016-04-01  0.73
## 5 2016-05-01  0.78
## 6 2016-06-01  0.43
## 7 2016-07-01  0.51
## 8 2016-08-01  0.49
## 9 2016-09-01  0.29
## 10 2016-10-01  0.4
## # ... with 57 more rows
```

O pacote {nucleos}

Como pode ser observado, uma vez criado e instalado o pacote, é muito fácil utilizar as funcionalidades para calcular as medidas de núcleos de inflação. Essas facilidades só são possíveis graças ao **processo de desenvolvimento do pacote**, que pode ser resumido em algumas **etapas**:

- 1) Criar lista com definição de itens e subitens que compõem cada núcleo (classificações do IPCA);
- 2) Criar funções para coleta e tratamento de dados;
- 3) Criar funções para calcular cada núcleo, utilizando os dados e definições da composição dos mesmos.

O pacote {nucleos}

Na **primeira etapa** criamos duas listas, uma para cada classificação dos itens do IPCA definida pelo BCB, respeitando a diferença de composição do índice de acordo com as últimas duas POF/IBGE⁹.

Cada lista contém 9 vetores numéricos com os códigos dos subitens e itens provenientes do SIDRA/IBGE, seguindo as classificações do BCB abordadas acima. As listas se encontram nos arquivos .R dentro do projeto¹⁰:

- ipca_classes_2012.R
- ipca_classes_2020.R

⁹POF 2008-09 utilizada no IPCA desde 2012 e POF 2017-18 utilizada no IPCA desde 2020.

¹⁰Com o pacote instalado é possível consultar a documentação referente a esses dados, que estão em formato de dataset no pacote, basta digitar no Console: `help("ipca_classes_2012")` ou `help("ipca_classes_2020")`.

O pacote {nucleos}

Na **segunda etapa** criamos duas funções para coletar e tratar os dados do IPCA do sistema SIDRA/IBGE usando a API de dados:

Função	Arquivo	Descrição
nucleos::get_ipca()	get_ipca.R	Possibilita a coleta e tratamento das tabelas 1419, 7060 do SIDRA/IBGE, referente aos dados desagregados do IPCA.
nucleos::group_desc()	group_desc.R	Possibilita criar duas colunas adicionais no output da função anterior com informações de código SNIPC e agrupamento (subitem, item, etc.).

O pacote {nucleos}

Por fim, na **terceira etapa** criamos as 4 funções para cálculos dos núcleos de inflação nos seguintes arquivos:

- core_adhoc.R
- core_ma.R
- core_ms.R
- core_dp.R

Núcleos de inflação

Agora que você já sabe como construir os núcleos de inflação e está um passo mais próximo de construir melhores modelos para a inflação!

Importante: na próxima atualização da POF/IPCA¹¹ provavelmente este pacote precisará ser atualizado, para respeitar a entrada/saída de subitens no índice de inflação e suas consequências nas classificações utilizadas nos núcleos.

¹¹A periodicidade da POF é, em princípio, quinquenal.

Curso de Análise de Conjuntura usando o R - Versão 5.0

Essa aula que vocês acabaram de ver faz parte do nosso Curso de Análise de Conjuntura usando o R.

Esse tipo de conhecimento, focado em habilidades, infelizmente, não é ensinado em graduações ou mesmo especializações, que cobram entre R\$ 10 mil e R\$ 30 mil por matrícula!

Esse gap foi **resolvido** aqui na Análise Macro com a criação do nosso Curso de Análise de Conjuntura usando o R!

Curso de Análise de Conjuntura usando o R - Versão 5.0

Ao longo dos últimos 5 anos, esse Curso foi frequentado por quase 400 alunos, dentre os quais:

- Estudantes de graduação e pós-graduação das principais faculdades do país;
- Professores das mesmas universidades;
- Profissionais de bancos, consultorias, institutos de pesquisas e grandes empresas. Em particular, temos entre nossos alunos Economistas-Chefes.

Curso de Análise de Conjuntura usando o R - Versão 5.0

Para a 2ª Turma da versão 5.0 que abrimos inscrições agora no nosso site, as principais inovações e benefícios serão:

- 30 Aulas Gravadas
- Certificado de 90 horas
- Plataforma tira-dúvidas exclusiva
- Grupo fechado no Whatsapp para networking dos alunos
- 12 Monitorias Especiais via Zoom (Bônus 01)
- Curso de R para Análise de Dados (Bônus 02)
- Curso de Visualização de Dados com ggplot2 (Bônus 03)

Um Curso sem igual no mercado!

Curso de Análise de Conjuntura usando o R - Versão 5.0

Para a 2ª Turma dessa nova Versão, abriremos apenas 20 vagas!

Queremos dar a essas 20 pessoas o melhor treinamento em análise de conjuntura disponível no mercado!

Além disso, vamos dar um desconto de R\$ 1500 diretamente no Site. Assim que atingirmos esse número, o preço será elevado automaticamente.