

# Índice de desenvolvimento Bruto

Ebenézer Dorneles

31/10/2022

## Introdução

O Projeto tem como um dos objetivos principais elaborar um modelo para a criação do Índice Bruto de Desenvolvimento (IDB) para os municípios que constituem a Região Geográfica Intermediária de Chapecó.

As informações foram coletadas em plataformas digitais com dados abertos. São estes: DATASUS, IBGE Cidades, SIDRA, IPEADATA e Atlas Brasil. Todos os dados são oriundos do Instituto Brasileiro de Geografia e Estatística entre os períodos do Censo 2000 e Censo 2010.

Váriáveis	nome
Densidade Demográfica	x1
% de população urbana	x2
% de pessoas em domicílios com energia elétrica 2010	x3
% da população em domicílios com água encanada 2010	x4
Nº Estabelecimentos de saúde	x5
Nº de leitos em estabelecimentos de saúde	x6
Taxa de analfabetismo - 15 anos ou mais de idade 2010	x7
População economicamente ativa de 18 anos ou mais de idade 2010	x8
Taxa de desocupação - 18 anos ou mais de idade 2010	x9
PIB percapta	x10
Receitas	x11
VAB Serviços	x12
VAB Indústria	x13
VAB Agropecuária	x14
Mortalidade infantil 2010	x15
IDMH	x16
IG	x17
% de pessoas em domicílios urbanos com coleta de lixo 2010	x18

## Análise Fatorial (AF)

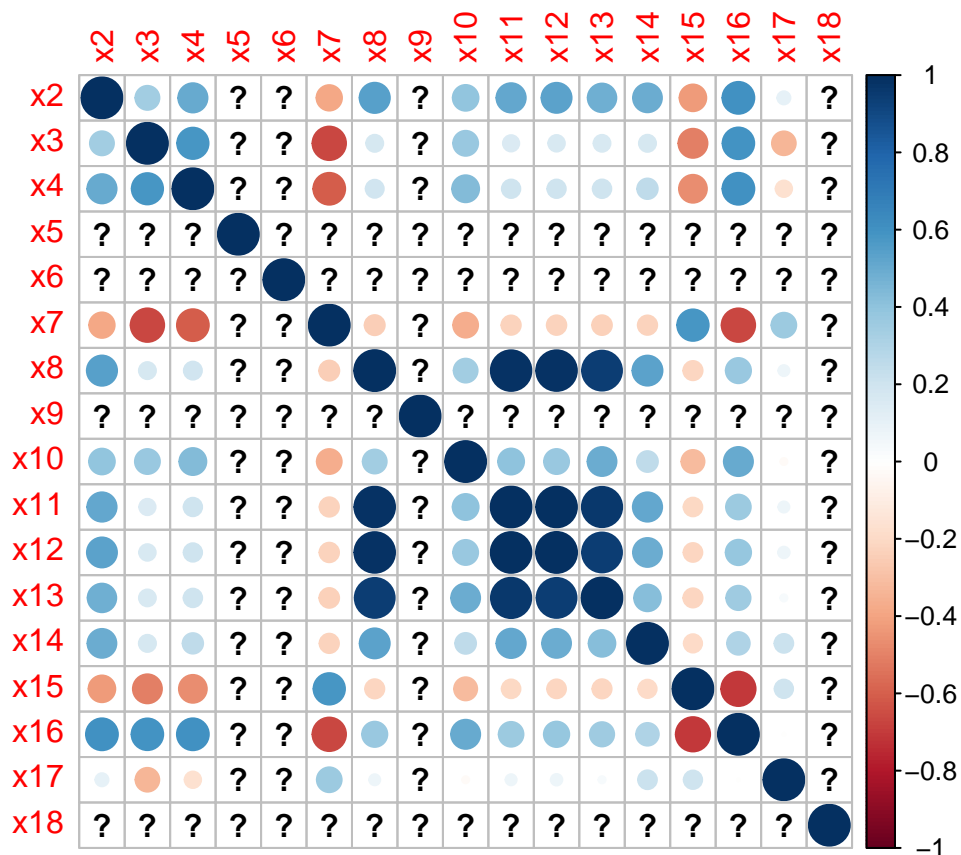
A Análise Fatorial de acordo Hair et al.(2006), é uma série de técnicas em um processo estatístico multivariado tornando possível a observação de variáveis conjuntas que apresentam a mesma estrutura subjacente. O método reduz o conjunto de dados em fatores ou dimensões que resumem as variáveis observadas. Um fator é a combinação linear das variáveis, com um peso que determina o quanto a variável contribui para o conjunto de dados observado. Antes de criarmos os fatores, precisamos verificar o quão adequado o modelo AF é para o conjunto de dados. Nesse processo utilizaremos os testes de correlação de Pearson, de Esfericidade de Bartlett e o teste de Kaiser-Meyer-Olkin (KMO). Os testes foram realizados no Rstudio em linguagem R com os pacotes GParotation, psych e corrplot.

## Teste de correlação de Pearson

Para determinar a matriz de correlação no conjunto de dados utilizaremos o método de correlação de Pearson que mede a relação linear entre duas variáveis quantitativas. Esse método correlaciona as variáveis pela variância compartilhada. Sendo assim, em um modelo de correlação linear supõe-se que o crescimento e o decrescimento de uma variável afete outra em mesmo impacto. Portanto, a correlação de Pearson exige um compartilhamento de variância e que essa variação seja distribuída linearmente (FIGUEIREDO FILHO E SILVA JÚNIOR, 2009). O coeficiente varia entre 1 (relação positiva) e -1 (relação negativa). Sendo o 0 a indicação de que não há correlação. No gráfico abaixo vemos a correlação das variáveis, os círculos em azul forte indicam uma relação próxima a 1 enquanto círculos na cor vermelha indicam uma relação próxima a -1.

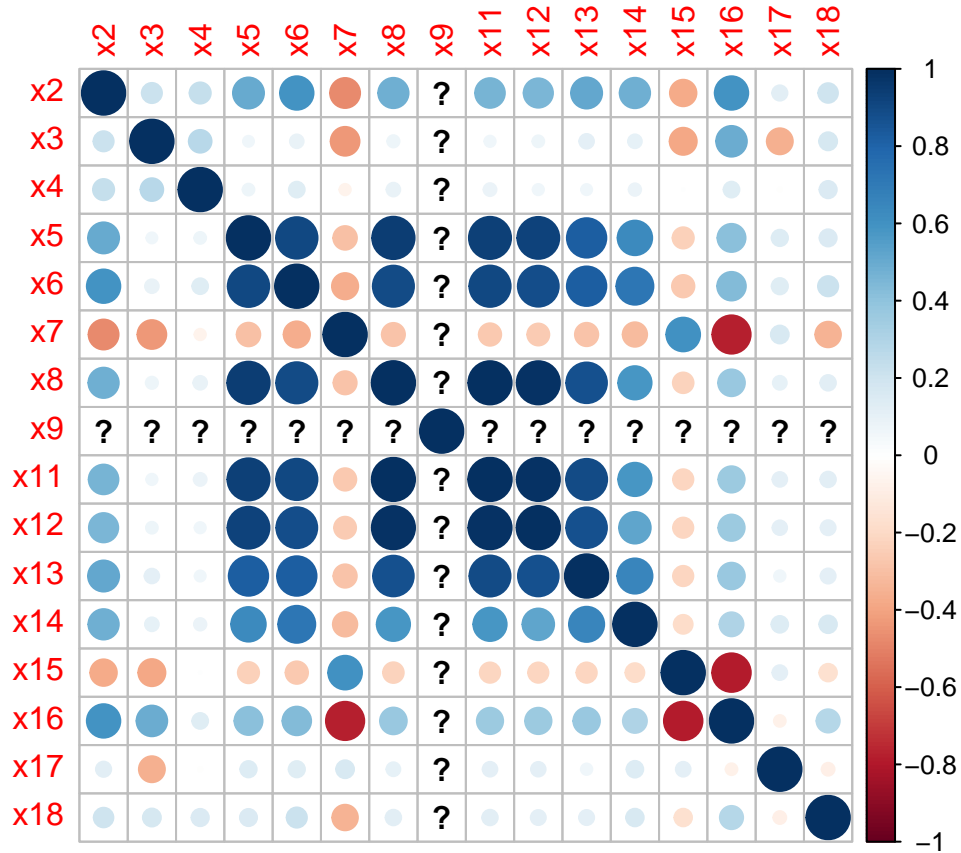
## Teste de correlação para 2000

Destarte nota-se que para o ano 2000, além das variáveis x1 e x10, as variáveis x5, x6, x9 e x18 não possuem relação nenhuma com as demais variáveis. A hipótese é que esses dados possuam valores faltantes para algumas observações, impossibilitando o cálculo de correlação. Para evitar que essas variáveis influenciem os testes seguintes será retirado esse conjunto dos dados para a análise.



## Teste de correlação para 2010

Para o ano de 2010 a correlação entre as variáveis aumentou positivamente, porém a variável x9 ainda demonstra uma incoerência em seus valores que impossibilita a efetivação do cálculo de correlação.



### Teste de Esfericidade de Bartlett

O teste de Esfericidade de Bartlett requer que as variáveis tenham distribuição normal. Sendo assim o teste verifica a correlação dos dados com a hipótese nula:  $H_0$  (matriz identidade), indicando que não há correlação entre as variáveis. Rejeitando a hipótese nula, temos que as variáveis se correlacionam entre si. Segundo Figueiredo filho e Silva Júnior (2010), o valor obtido com o teste de Esfericidade de Bartlett deve ser estatisticamente significativo ( $P\text{-valor} < 0,05$ ).

Table 2: Teste para o ano 2000

p-value
0.0059751

Table 3: Teste para o ano 2010

p-value
0.0200491

### Teste de Kaiser-Meyer-Olkin (KMO)

O teste KMO verifica a adequação do conjunto de dados com o modelo da Análise Fatorial. Segundo Mingoti (2007), quando as correlações parciais estão próximas a zero, o coeficiente KMO está próximo de um, o que indica que o modelo está bem ajustado com os dados. O mínimo para que o modelo fatorial seja adequado

varia de autor para autor. Destarte, aceita-se que o modelo seja adequado quando  $KMO > 0,5$ . Caso o valor não esteja ajustado, a correção nos dados amostrais deverá ser feita através da exclusão de variáveis dentre as avaliadas, ou então, a inclusão de novas variáveis para melhorar a adequação do modelo.

Table 4: Teste para o ano 2000

KMO geral
0.7925168

Table 5: Teste para o ano 2010

KMO geral
0.7811909

### **Análise Fatorial Exploratória**

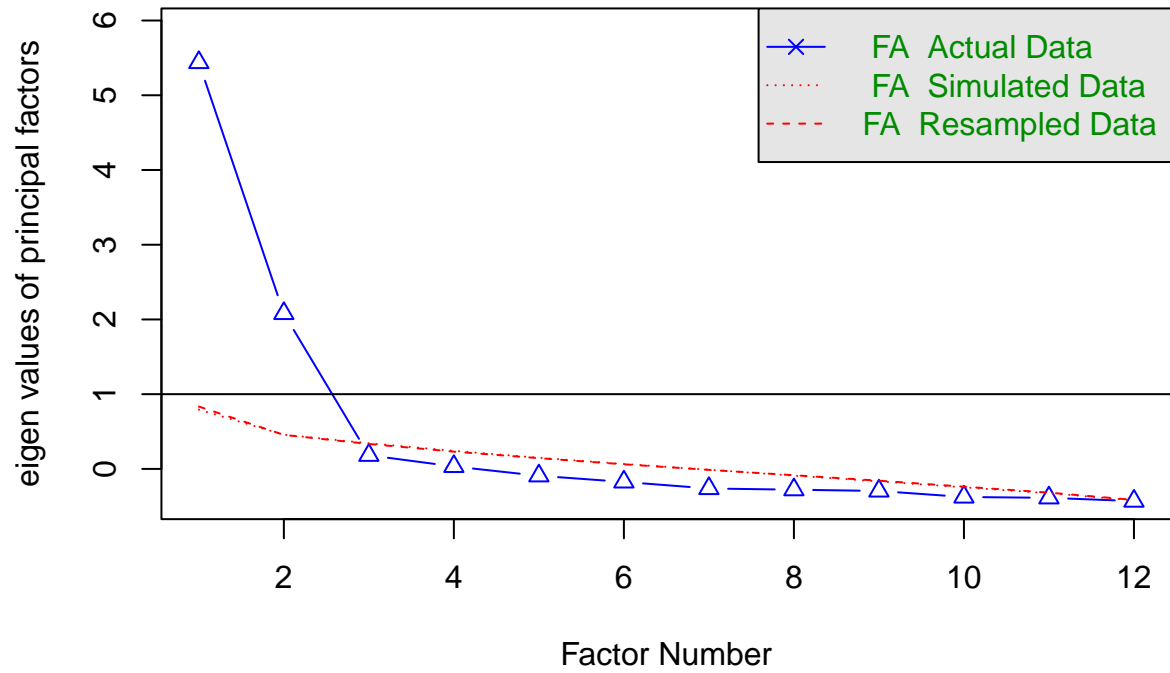
No decorrer dos testes de correlação, Esfericidade de Bartlett e Kaiser-Meyer-Olkin foi preciso realizar ajustes no conjunto de variáveis para que o modelo seja o mais adequado possível, pois a análise se baseia na matriz de correlação/covariância para elaboração dos fatores. A partir desses testes podemos adequar e melhorar o conjunto de dados para o modelo de análise fatorial.

Portanto, precisamos determinar as variáveis latentes do nosso conjunto, sendo assim utilizaremos a técnica de análise paralela do pacote physis. Esta função realiza interações entre os principais autovalores dos fatores com o peso de cada fator determinado pela variância do conjunto de dados. Por meio dessas interações é possível determinar o número adequado de fatores para utilização. Além deste teste, utilizaremos o critério empregado por De Oliveira e Da Silva (2017), em que consideramos a escolha dos fatores aqueles que possuem um autovalor  $> 1$ .

### **Análise Exploratória para 2000**

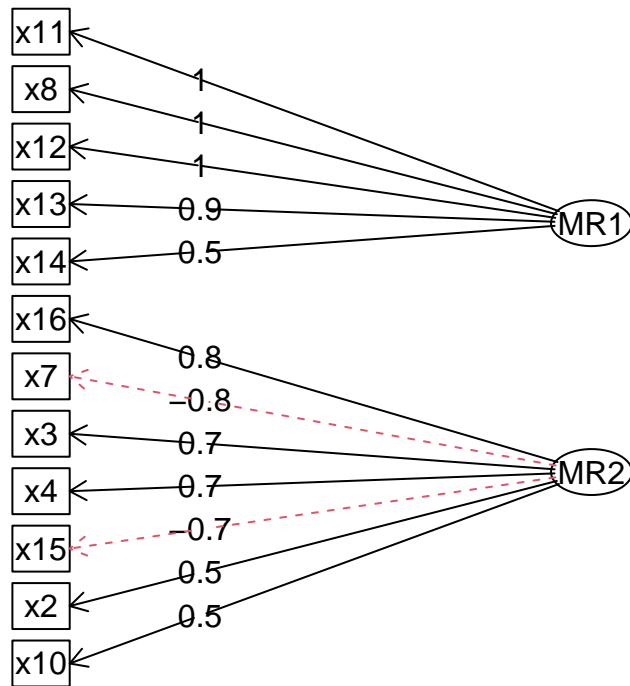
Para o ano 2000, temos dois fatores significativos para análise com um autovalor maior que um. Além disso, para cada fator existem as seguintes correlações.

## Análise Paralela



## Parallel analysis suggests that the number of factors = 2 and the number of components = NA

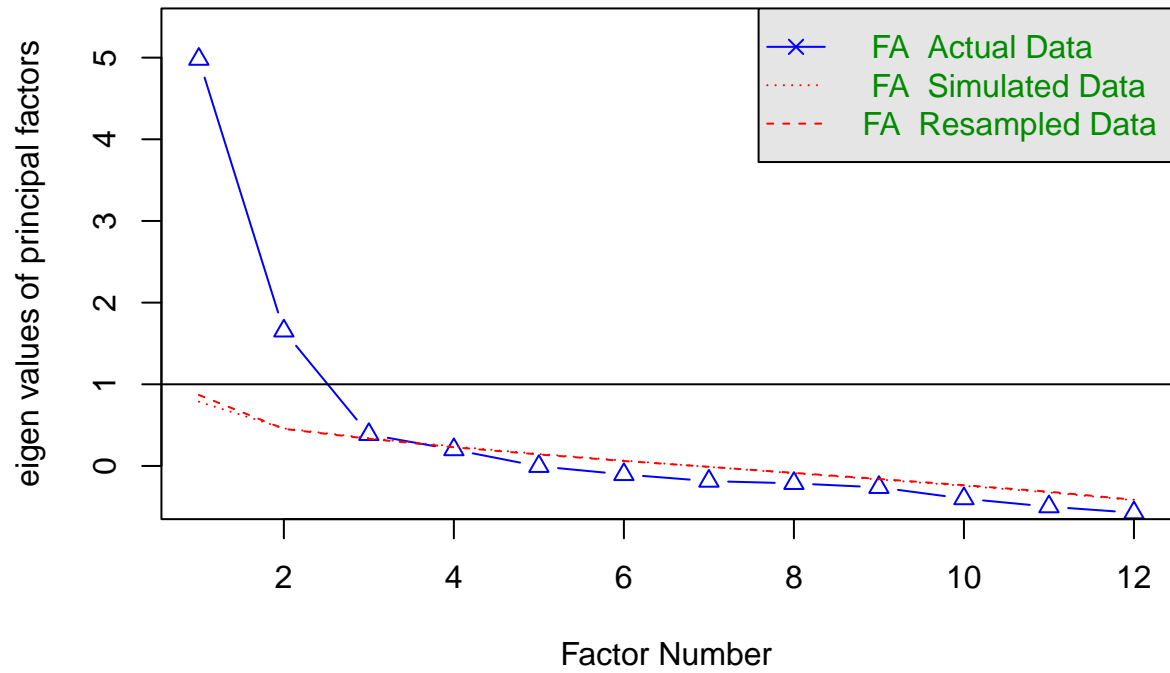
## Factor Analysis



### Análise Exploratória para 2010

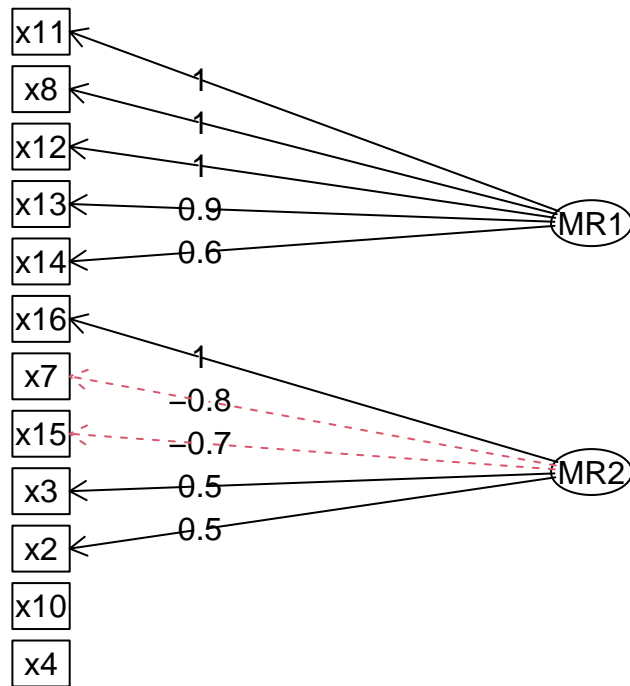
Para o ano de 2010, também é indicado a utilização de dois fatores, ambos com autovalores maiores que um, possuindo as seguintes correlações.

## Análise Paralela



## Parallel analysis suggests that the number of factors = 2 and the number of components = NA

## Factor Analysis



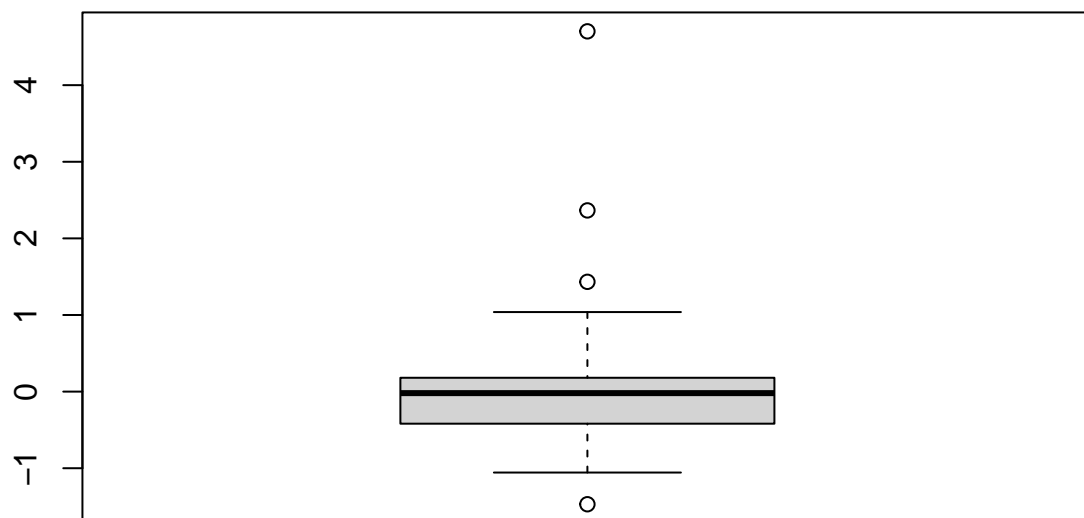
### Índice de Desenvolvimento Bruto

Ao determinarmos os fatores e os seus principais autovalores podemos determinar os scores fatoriais para cada observação do nosso conjunto de dados pelo método de regressão (Field, 2009). Destarte, obtemos o Índice de Desenvolvimento Bruto através da média ponderada pela raiz característica de cada fator.

### Índice de desenvolvimento Bruto para 2000

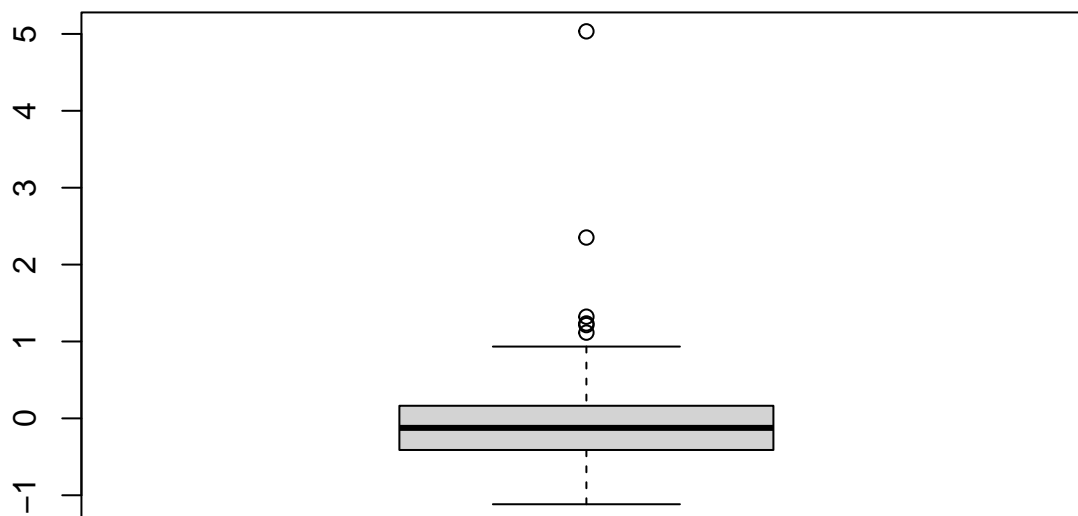
Dados descritivos
Min. :-1.4701
1st Qu.: -0.4168
Median :-0.0197
Mean : 0.0000
3rd Qu.: 0.1803
Max. : 4.7023





### Índice de desenvolvimento Bruto para 2010

Dados descritivos
Min. :-1.1167
1st Qu.: -0.4095
Median :-0.1237
Mean : 0.0000
3rd Qu.: 0.1616
Max. : 5.0335



## Referências

1. DE OLIVEIRA, Marines Rute; DA SILVA, Gerson Henrique. Análise espacial do desenvolvimento econômico dos municípios do oeste do Paraná. **Revista Capital Científico-Eletrônica (RCC)-ISSN 2177-4153**, v. 15, n. 2, p. 62-78, 2017.
2. HAIR, Jr; BLACK, W. C; BABIN, B. J; ANDERSON, R. E e TATHAM, R. L. **Multivariate Data Analysis**. 6ª edição. Upper Saddle River, NJ: Pearson Prentice Hall, 2006.
3. MINGOTI, S. **Análise de dados através de métodos de estatística multivariada: uma Abordagem Aplicada**. Editora UFMG, 2007.
4. FIGUEIREDO FILHO, Dalson Britto; SILVA JUNIOR, José Alexandre. **Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r)**. Revista Política Hoje, v. 18, n. 1, p. 115-146, 2009.
5. FIGUEIREDO FILHO, Dalson Brito; SILVA JÚNIOR, José Alexandre da. **Visão além do alcance: uma introdução à análise fatorial**. Opinião pública, v. 16, n. 1, p. 160-185, 2010.
6. REVELLE, W. (2020) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Version = 2.0.12,.
7. BERNAARDS, Coen A. and JENNRICH, Robert I. (2005) Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis, Educational and Psychological Measurement: 65, 676-696. <<http://www.stat.ucla.edu/research/gpa>>
8. TAIYUN wei and VILIAM Simko (2017). R package “corrplot”: Visualization of a Correlation Matrix (Version 0.84). Available from <https://github.com/taiyun/corrplot>