Escola Politécnica de Pernambuco

Especialização em Ciência de Dados e Analytics

Estatística Computacional

Aula 4.1 – Análise de Conglomerados e Fatorial – PARTE II

Prof. Dr. Rodrigo Lins Rodrigues

rodrigo.linsrodrigues@ufrpe.br

O que veremos nesta aula ?

- Entender **quando** é necessário **utilizar análise fatorial**;
- Entender as **premissas** para a utilização da análise fatorial;
- Diferenciar A.F de outras técnicas multivariadas;
- Entender os principais **métodos de rotação**;
- Determinar o **número de fatores** a serem extraídos;
- Nomear fatores;
- Saber o que é **carga fatorial**;
- Aplicações computacionais.

Análise Fatorial

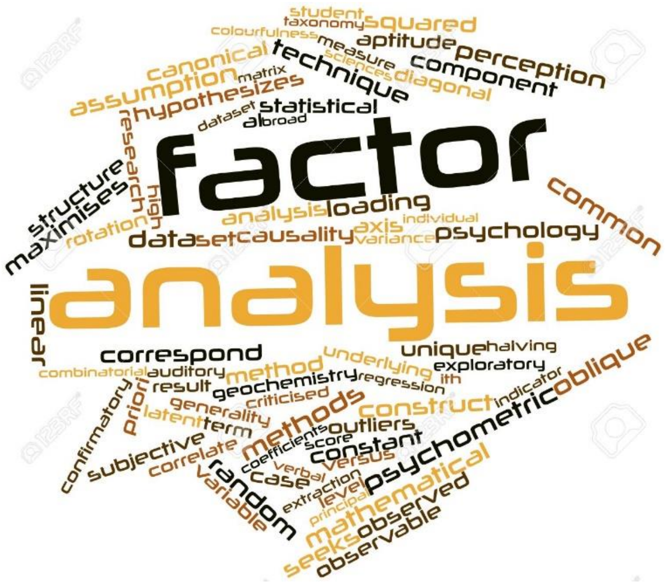
Análise Fatorial

- É uma **técnica multivariada** que busca identificar **relações entre variáveis** inter-relacionadas;

- Busca identificar **fatores comuns**;

- Busca representar um conjunto de variáveis por meio de um **número menor de fatores**.

- As variáveis são agrupadas em função de suas correlações;

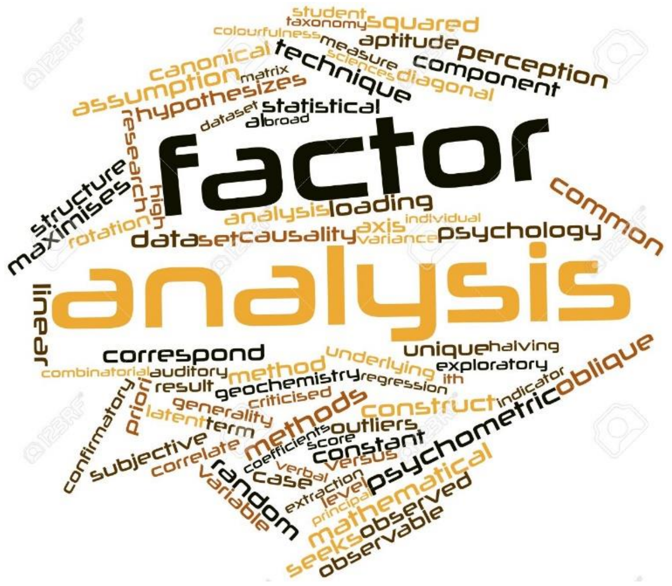


Análise Fatorial

- A maior vantagem da A.F é **simplificar ou reduzir** um número grande dados, por intermédio da determinação de fatores.;

- Possibilita ao pesquisador a criação de indicadores **inicialmente não observáveis** compostos por **agrupamento de variáveis**;

- A possibilidade de utilização são inúmeras.



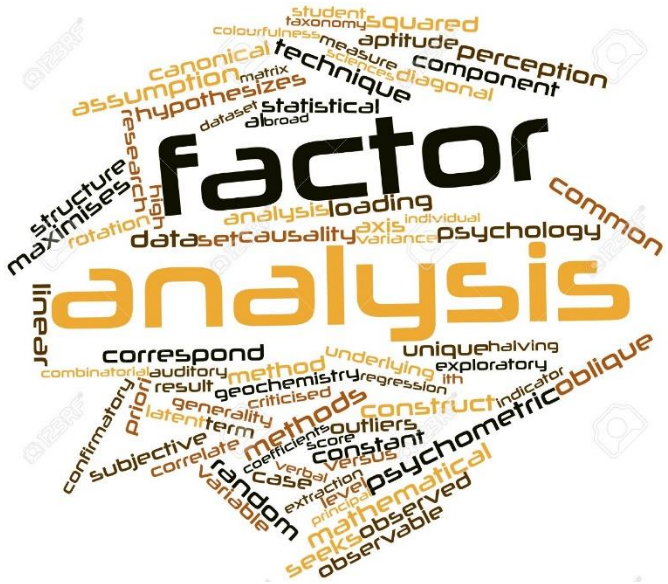
Análise Fatorial

- Um fator é uma **combinação linear** das variáveis originais;

- As suposições para A.F são:

  ✓**Normalidade e linearidade:** Deve-se ter a garantia de normalidade multivariada;

  ✓**Matriz de correlações com valores significativos:** número substancial de correlações superiores a 0,30

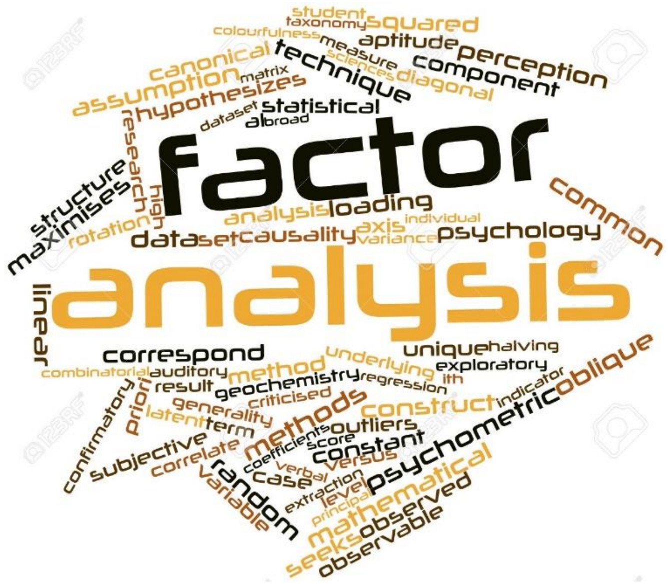


Análise Fatorial

- A aplicação de A.F deve ser realizada em amostras com tamanhos igual ou superior a 100 observações;

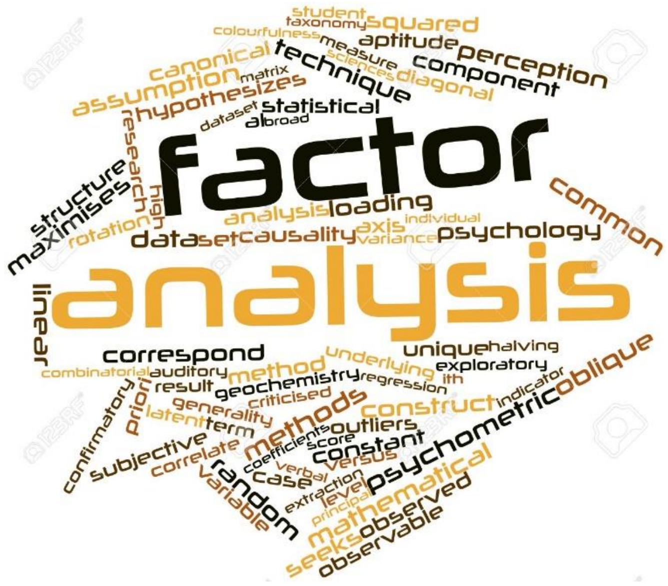
- A análise fatorial pode ser:

  ✓Exploratória;
  ✓Confirmatória.



Análise Fatorial

- **Análise Fatorial Confirmatória:**
  - ✓O pesquisador possui conhecimento prévio sobre o relacionamento das variáveis e fatores;
  - ✓Busca confirmar uma estrutura previamente estabelecida pelo pesquisador.



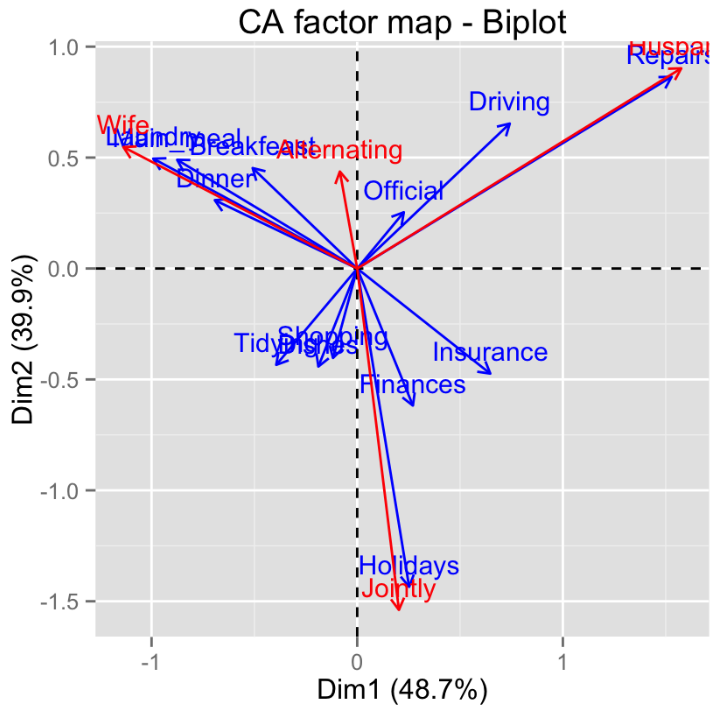
- **Análise Fatorial Exploratória:**
  - ✓Neste caso o pesquisador tem pouco ou nenhum conhecimento prévio sobre a estrutura dos fatores;
  - ✓Abordaremos apenas a A.F Exploratória.

Modelagem da Análise Fatorial

Modelagem da Análise Fatorial

Os passos para a modelagem de A.F.E são:

1. Análise da **matriz de correlações** e adequação da utilização da A.F;
2. Extração dos **fatores iniciais** e determinação do **número de fatores**;
3. **Rotação** dos fatores;
4. **Interpretação** dos fatores.



Modelagem da Análise Fatorial

1. Análise da Matriz de Correlações:

- ✓A A.F é baseada nas **correlações entre as variáveis**;
- ✓O primeiro passo é **construir** e **examinar** a matriz de correlações;
- ✓Verificar se existem **valores significativos** para justificar o uso da técnica;
- ✓Caso a matriz seja composta por valores baixos de correlação não será ideal a A.F;
- ✓É necessário verificar a **normalidade multivariada**.

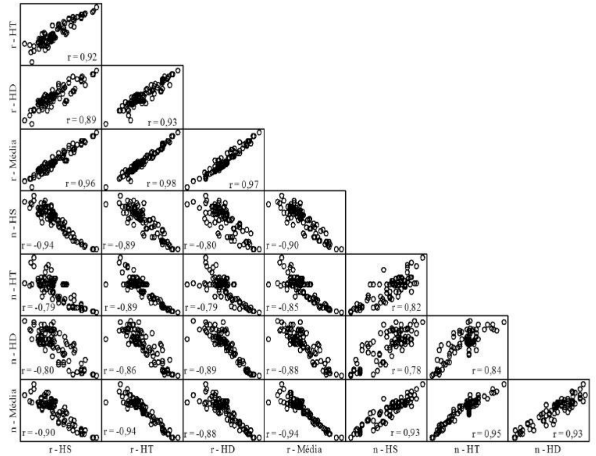


Figura 1. Diagramas de dispersão entre as 91 estimativas (14 caracteres tomados dois a dois) do coeficiente de correlação linear de Pearson (r) e do tamanho de amostra, em número de plantas (n), para a estimação de r para a amplitude do intervalo de confiança de "bootstrap" de 95%, igual a 0,30, referentes aos híbridos simples (HS), triplo (HT), duplo (HD) e média dos três híbridos de milho. Valor de r superior a |0,20| é significativo a 5% de probabilidade, pelo teste t, com 89 graus de liberdade.

Modelagem da Análise Fatorial

1.1 KMO e Teste de Esfericidade de Bartlett:

- ✓**Verificar a adequação** da matriz de correlação a A.F;
- ✓Avalia a hipótese de que a matriz das correlações pode ser uma **matriz identidade** com determinando igual a 1:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- ✓Se a matriz de correlação for igual à matriz de identidade, significa que as inter-relações entre as variáveis são iguais a 0;

Modelagem da Análise Fatorial

1.1 KMO e Teste de Esfericidade de Bartlett:

✓É usada a estatística KMO (Kaiser-Mayer-Olkin) para realizar o teste:

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum a_{ij}^2}$$

Em que:

r_{ij}= coeficiente de correlação entre variáveis;

a_{ij} = coeficiente de correlação parcial;

Modelagem da Análise Fatorial

1.1 KMO e Teste de Esfericidade de Bartlett:

- ✓ Os valores de KMO variam entre 0 e 1. Valor próximo de 0 indica que a análise fatorial pode não ser adequada para o estudo;
- ✓ Pois existe uma correlação fraca entre as variáveis;
- ✓ A tabela a baixo mostra os valores de referência:

KMO	Adequação a Análise Fatorial
0,9 |-- 1,0	Muito Boa
0,8 |-- 0,9	Boa
0,7 |-- 0,8	Média
0,6 |-- 0,7	Razoável
0,5 |-- 0,6	Ruim
< 0,5	Inaceitável

Modelagem da Análise Fatorial

2. Extração dos Fatores Iniciais:

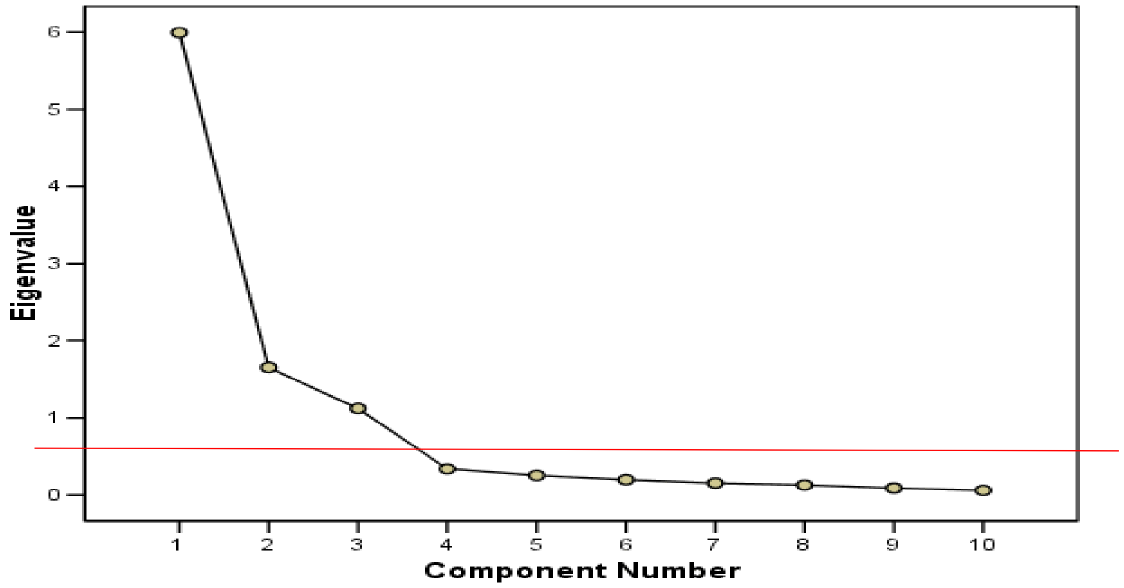
- ✓Nessa etapa é determinado o número de fatores;
- ✓O método extrai, inicialmente, a combinação linear que explica a maior parte da variância dos dados;
- ✓Na sequencia busca-se uma nova combinação (fator);
- ✓A priori o pesquisador informa o número de fatores de interesse.
- ✓Existem critérios que podem auxiliar na decisão do número de fatores:
  - Critério da raiz latente (critério de Kaiser);
  - Critério de percentagem de variância;
  - **Critério do gráfico Scree plot (vamos utilizar este);**
  - ...

Modelagem da Análise Fatorial

2. Extração dos Fatores Iniciais:

2.1 Gráfico Scree plot

- ✓ O ponto a partir do gráfico passa a se tornar “mais horizontal” reflete um indicativo do número máximo de fatores a serem extraídos:



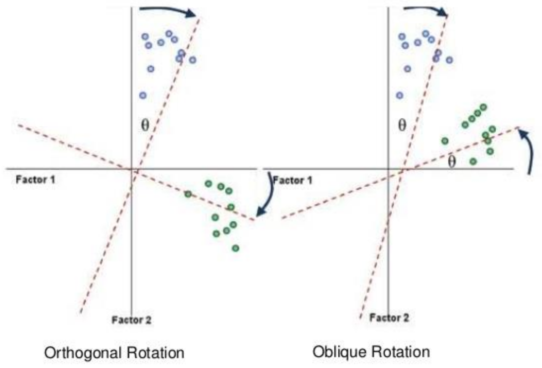
Modelagem da Análise Fatorial

3. Rotação dos Fatores:

- ✓Normalmente, os fatores produzidos na fase de extração **nem sempre** são **facilmente interpretados**;

- ✓A rotação de fatores tem como objetivo a transformação dos coeficientes dos componentes possibilitando **uma melhor descriminação**;

- ✓Alguns métodos de rotação são:
  - ✓**Varimax (utilizaremos este);**
  - ✓Quartimax
  - ✓Equamax.



Modelagem da Análise Fatorial

4. Interpretação dos Fatores:

- ✓A última etapa da técnica de A.F refere-se a **interpretação e nomeação dos fatores**;

- ✓É por meio das **cargas fatoriais** que fazem-se as interpretações;

- ✓De acordo com o valor das cargas é que identificamos em qual **fator cada variável pertence**.

Prática Computacional no SPSS

Prática Computacional - SPSS

- **Exemplo prático no SPSS**
  - ✓Um analista de mercado quer estudar as **relações estruturais** entre **quatro indicadores** financeiros provenientes de 45 empresas;

1. Código da empresa (Cod_Emp);
2. Prazo médio de recebimento das vendas (PMRV, em dias);
3. Endividamento (em %);
4. Vendas (em R$);
5. Margem líquida das vendas (em %)

Prática Computacional - SPSS

- **Exemplo prático no SPSS**
  - ✓A tabela de análise abaixo:

Fatorial.sav [Conjunto_de_dados1] - IBM SPSS Statistics Editor de dados

Arquivo Editar Visualizar Dados Transformar Analisar Marketing direto Gráficos Utilitários Janela

	Cod_Emp	PMRV	Endividamento	Vendas	Margem_liquida	var
1	1	41,50	25,57	4244,69	15,30	
2	2	97,37	36,27	5831,50	17,98	
3	3	82,39	38,63	4629,89	16,59	
4	4	96,30	22,26	4396,63	17,98	
5	5	99,51	15,09	4770,06	18,19	
6	6	96,30	48,36	6952,86	16,91	
7	7	95,23	33,81	5342,51	17,12	
8	8	96,30	26,64	6796,64	17,66	
9	9	99,80	33,71	9640,70	18,19	
10	10	62,06	16,26	3412,23	16,91	
11	11	79,18	18,30	3719,32	17,12	

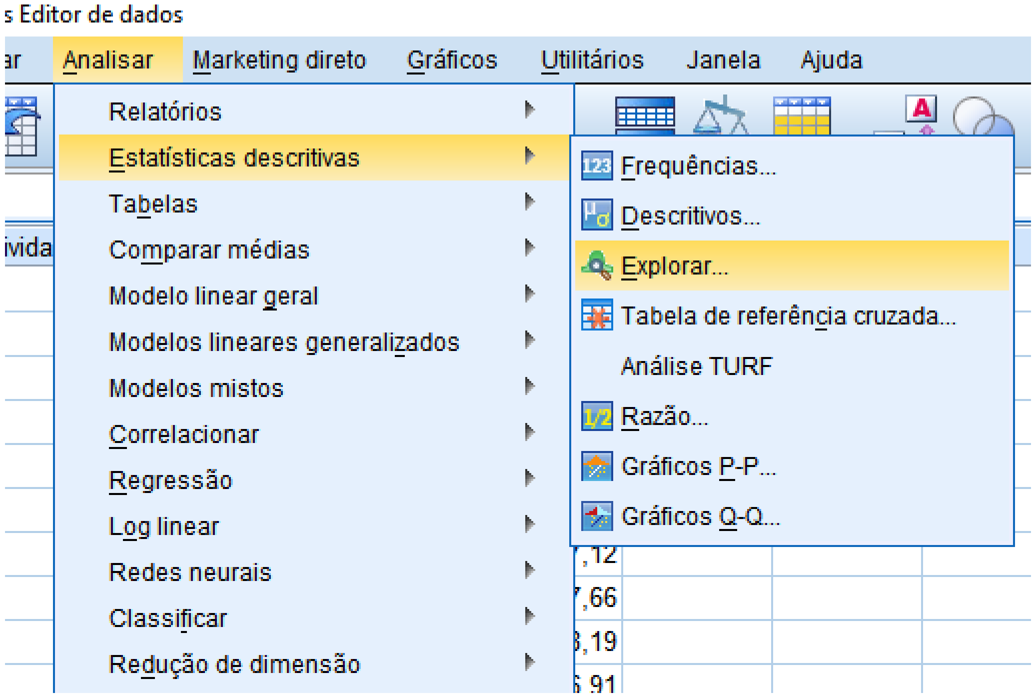
Prática Computacional - SPSS

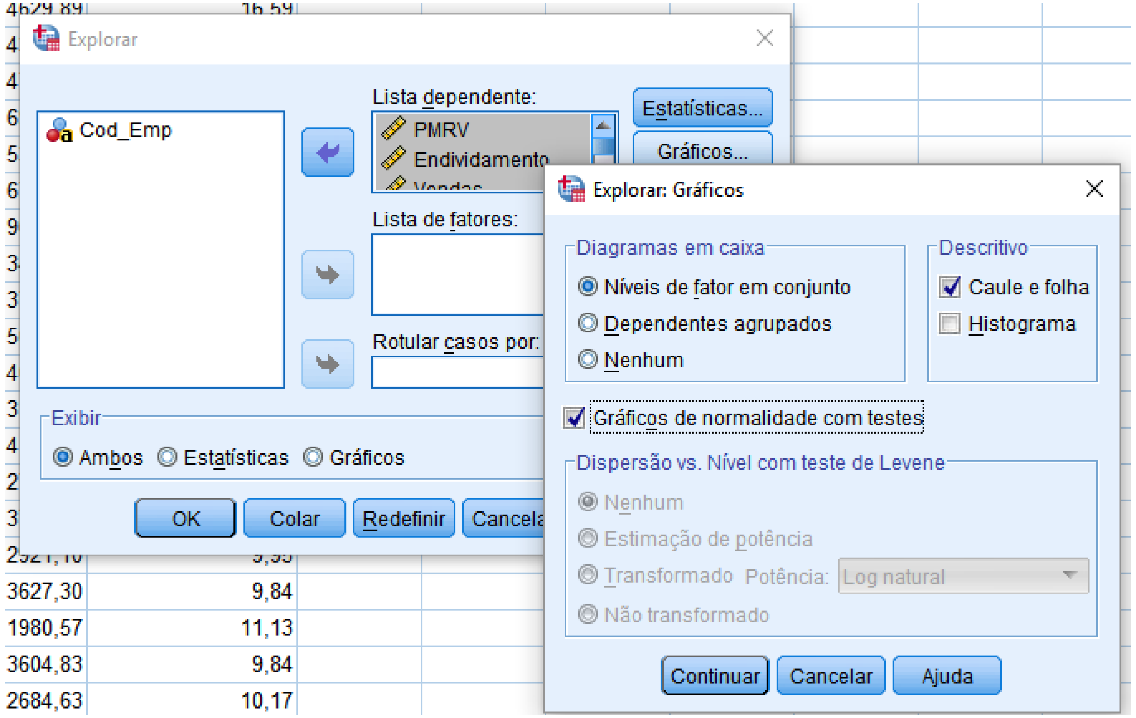
- **Exemplo prático no SPSS**
  - ✓Obs: O software SPSS não faz análise de normalidade multivariada, apenas teste de normalidade em uma variável por vez;

"Apesar da normalidade univariada não garantir a normalidade multivariada, e todas as variáveis atendem a essa condição, então quaisquer desvios de normalidade multivariada geralmente são inúculos" (Hair, Anderson, 2005)

Prática Computacional - SPSS

- Testando a normalidade de cada variável individualmente





Prática Computacional - SPSS

- Testando a normalidade de cada variável individualmente

Testes de Normalidade

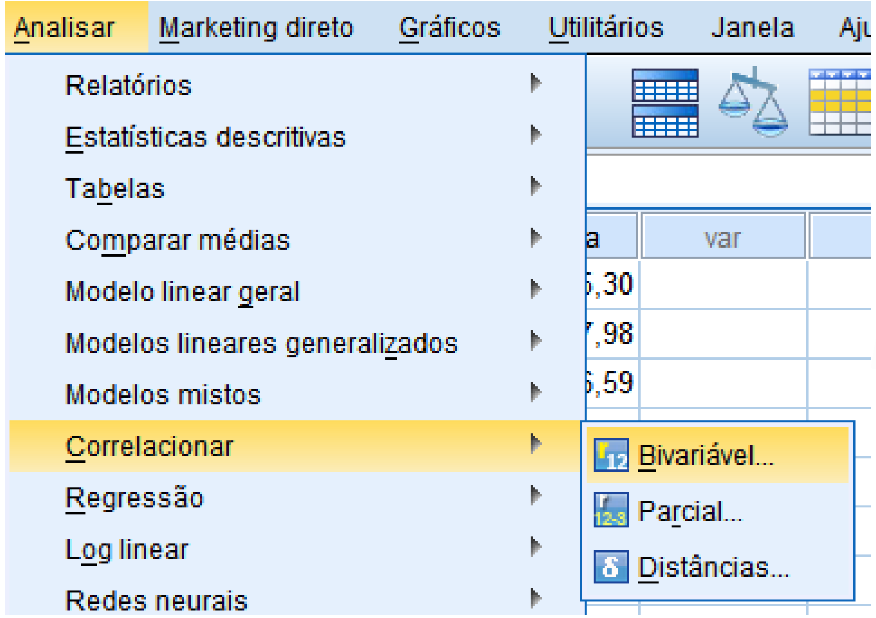
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estatística	df	Sig.	Estatística	df	Sig.
PMRV	,107	45	,200*	,914	45	,003
Endividamento	,120	45	,100	,940	45	,021
Vendas	,126	45	,073	,907	45	,002
Margem_liquida	,148	45	,015	,906	45	,001

*. Este é um limite inferior da significância verdadeira.

a. Correlação de Significância de Lilliefors

Prática Computacional - SPSS

- Gerando a matriz de correlação

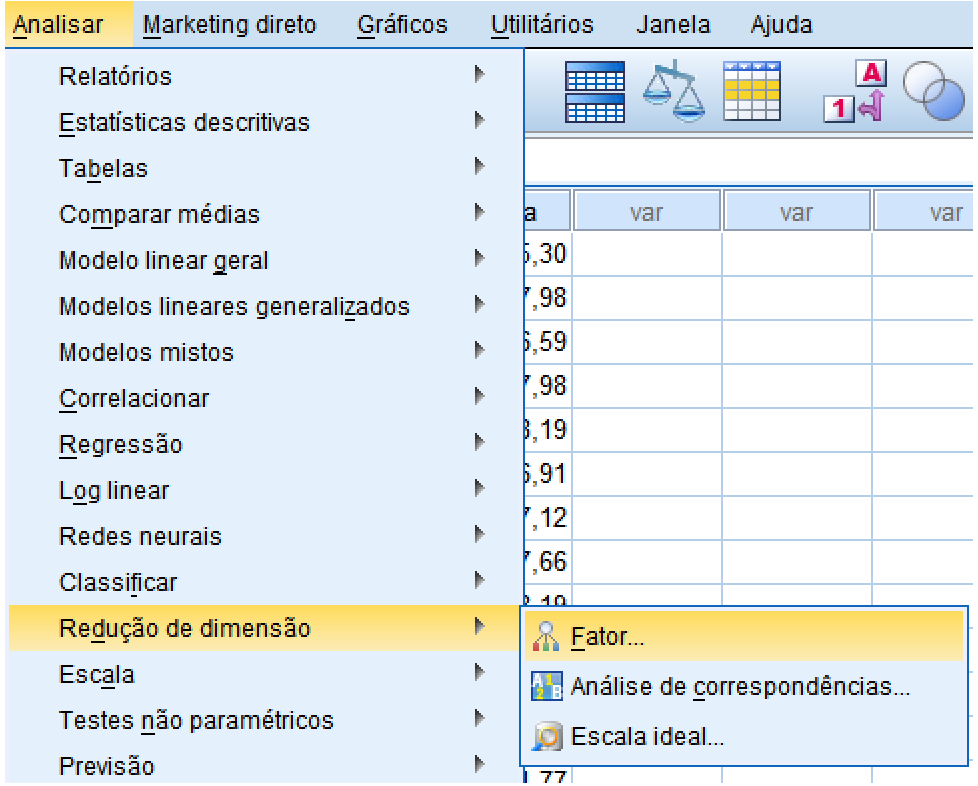


Correlações

		PMRV	Endividamento	Vendas	Margem_liquida
PMRV	Correlação de Pearson	1	,235	,625**	,598**
	Sig. (2 extremidades)		,121	,000	,000
	N	45	45	45	45
Endividamento	Correlação de Pearson	,235	1	,238	-,098
	Sig. (2 extremidades)	,121		,115	,523
	N	45	45	45	45
Vendas	Correlação de Pearson	,625**	,238	1	,580**
	Sig. (2 extremidades)	,000	,115		,000
	N	45	45	45	45
Margem_liquida	Correlação de Pearson	,598**	-,098	,580**	1
	Sig. (2 extremidades)	,000	,523	,000	
	N	45	45	45	45

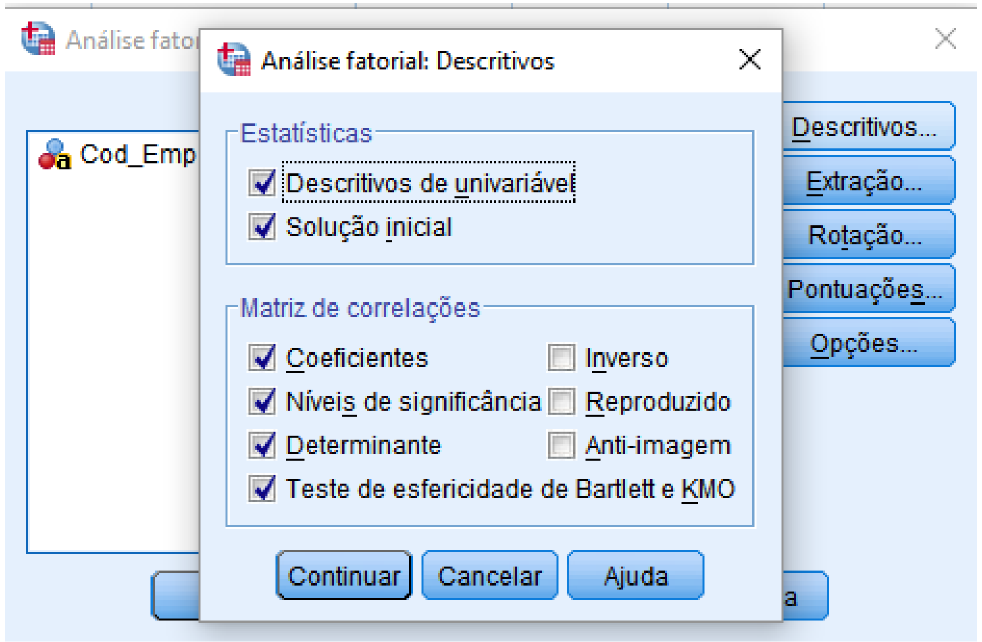
Prática Computacional - SPSS

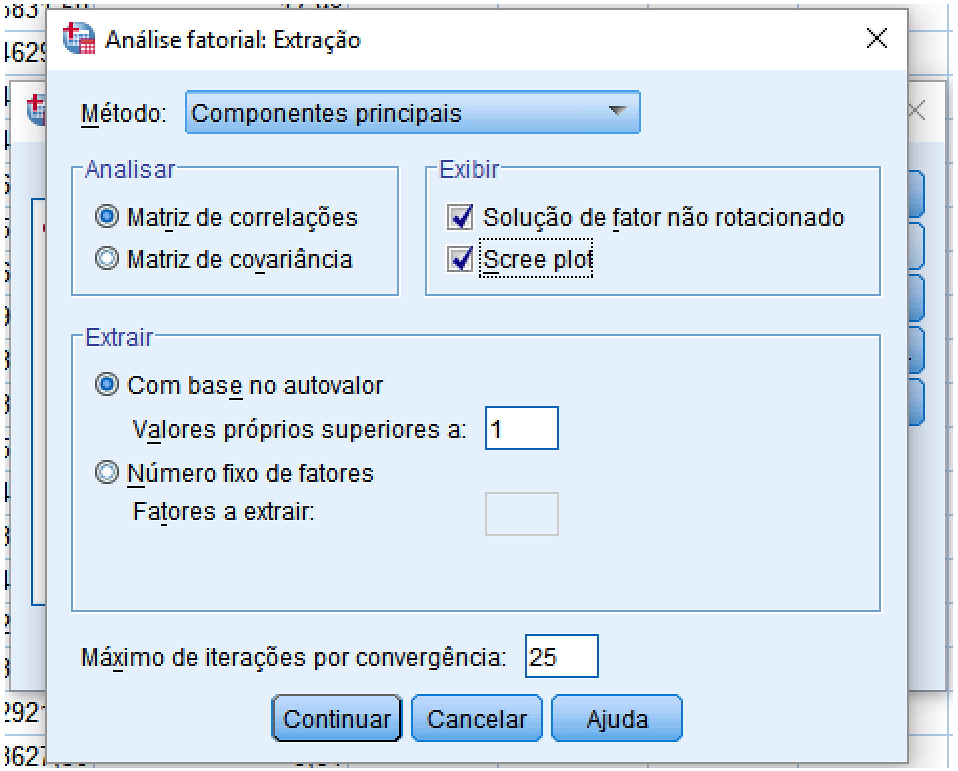
- Aplicando a Análise Fatorial – A.F



Prática Computacional - SPSS

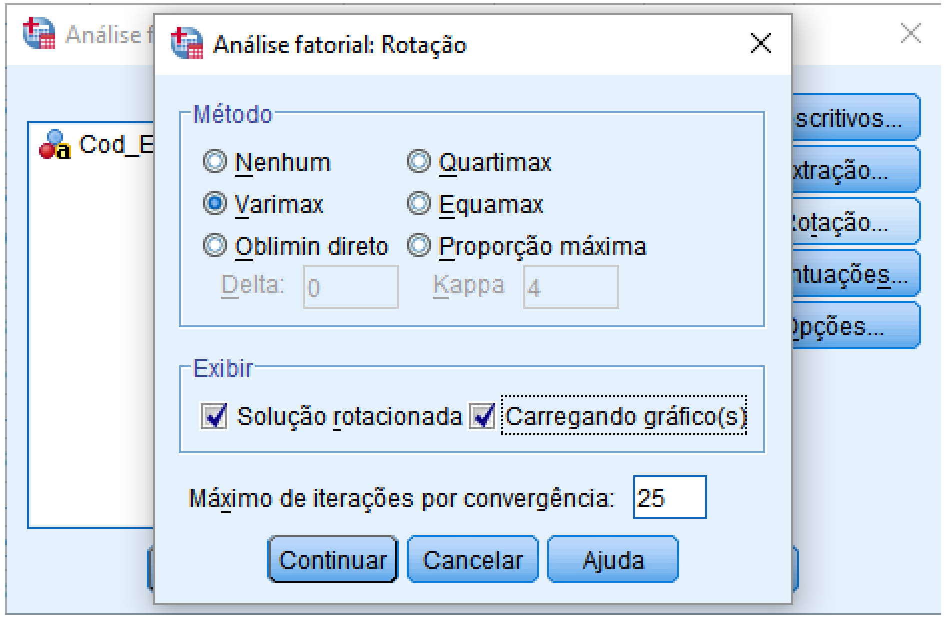
- Aplicando a Análise Fatorial – A.F





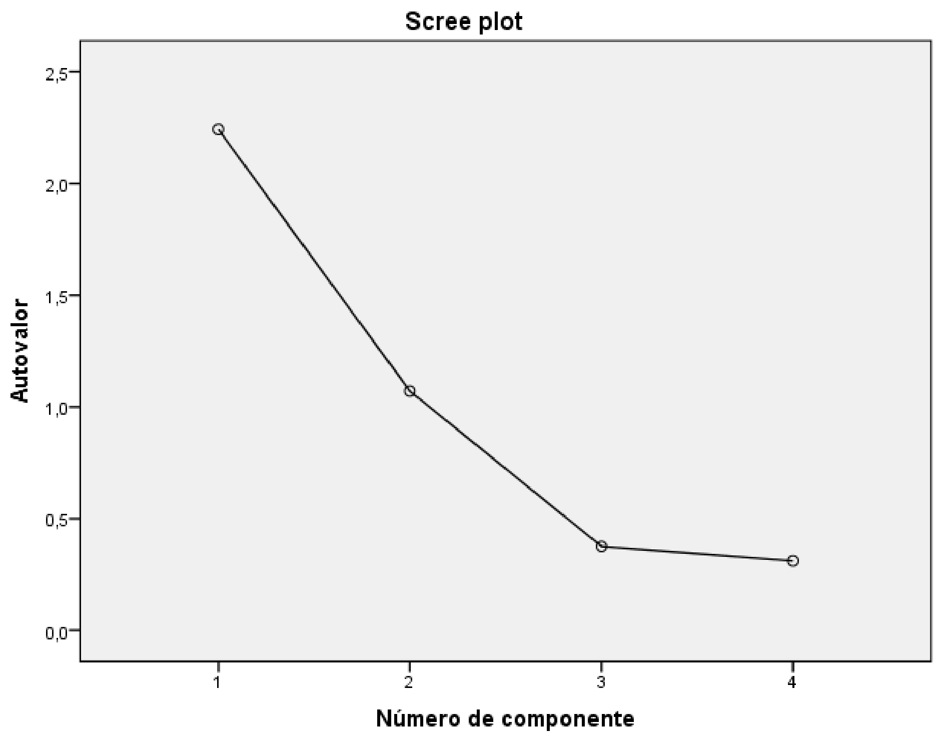
Prática Computacional - SPSS

- Aplicando a Análise Fatorial – A.F



Prática Computacional - SPSS

- Resultado da Aplicando a Análise Fatorial – A.F



Matriz de componente[a]

	Componente	
	1	2
PMRV	,876	,044
Endividamento	,269	,940
Vendas	,868	,061
Margem_liquida	,806	-,427

Matriz de componente rotativa[a]

	Componente	
	1	2
PMRV	,850	,215
Endividamento	,079	,974
Vendas	,840	,230
Margem_liquida	,874	-,261

Prática Computacional no R

Prática Computacional - R

- **Exemplo prático no R**
  - ✓Um analista de mercado quer estudar as **relações estruturais** entre **quatro indicadores** financeiros provenientes de 45 empresas;

1. Código da empresa (Cod_Emp);
2. Prazo médio de recebimento das vendas (PMRV, em dias);
3. Endividamento (em %);
4. Vendas (em R$);
5. Margem líquida das vendas (em %)

Prática Computacional - R

- **Exemplo prático no R**

```r
# Importando dados do spss
dados <- read.spss("Fatorial.sav", to.data.frame = TRUE)
dados <-dados[,2:5] #retirando a primeira coluna

# Calculando as estatisticas descritivas
sapply(dados, summary)
```

```
         PMRV Endividamento Vendas Margem_liquida
Min.     6.42         14.77   1981          8.453
1st Qu. 25.68         22.00   2921         10.160
Median  49.22         29.75   3719         13.050
Mean    53.13         31.71   3989         13.220
3rd Qu. 82.39         39.00   4770         16.910
Max.    99.80         69.44   9641         18.190
```

Prática Computacional - R

- **Matriz de Correlação**

```r
# Calculando a matriz de correlação
cor(dados)
```

```
                   PMRV Endividamento    Vendas Margem_liquida
PMRV           1.0000000    0.23454105 0.6253536     0.59834747
Endividamento  0.2345410    1.00000000 0.2384081    -0.09781364
Vendas         0.6253536    0.23840813 1.0000000     0.57970315
Margem_liquida 0.5983475   -0.09781364 0.5797031     1.00000000
```

Prática Computacional - R

- **Teste de esfericidade de Bartlett**

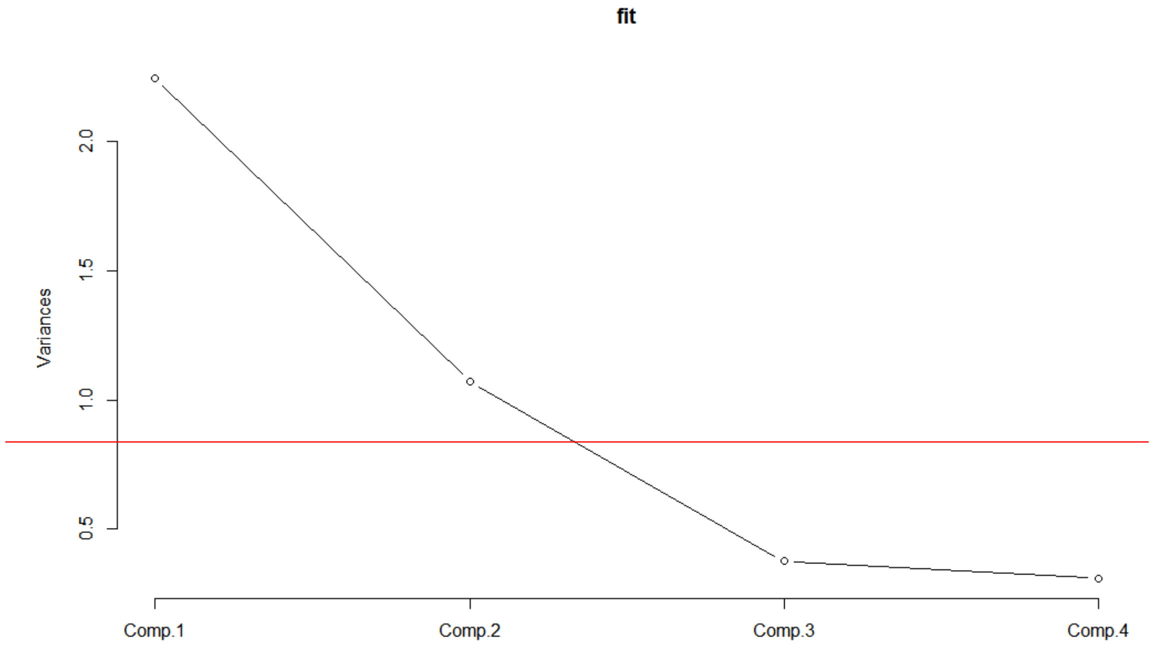
```r
# Verificando a adequação com o teste de esfericidade de Bartlett
# se o p-valor for menor do que 0,05 rejeita-se a hipotese h0 (matriz identidade)
cortest.bartlett(dados)
```

```
$chisq
[1] 53.16543

$p.value
[1] 1.086646e-09

$df
[1] 6
```

Prática Computacional - R

- **Verificando quantidade de fatores com ScreenPlot**

```r
# Gerando o screen plot para verificar a quantidade de fatores
fit<-princomp(dados, cor="TRUE")
plot(fit, type="line")
```



Prática Computacional - R

- **Extraindo fatores pelo método varimax**

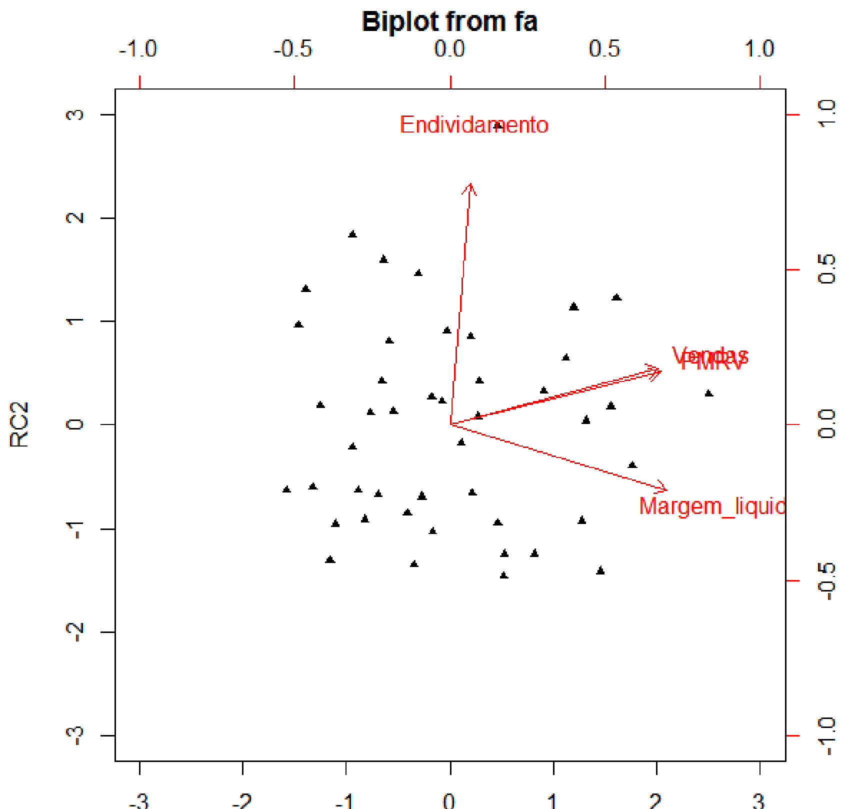
```
# Extraindo os fatores pelo método varimax
fit<-principal(dados, nfactors = 2, rotate = "varimax", scores = TRUE)
```

```
# Verificando cargas fatoriais para cada fator
load <- fit$loadings
load
```

```
Loadings:
                RC1    RC2
PMRV            0.850  0.215
Endividamento          0.974
Vendas          0.840  0.229
Margem_liquida  0.874 -0.261
```

Prática Computacional - R

- **Verificando as componentes graficamente**



Dúvidas ?

- **Contatos:**
  - ✓Email: rodrigo.linsrodrigues@ufrpe.br
  - ✓Facebook: /rodrigomuribec