

# ANÁLISE MULTIVARIADA DE DADOS – ANÁLISE FATORIAL

Andrezza Alves Pires<sup>1</sup>; Tarcísio da Costa Lobato<sup>2</sup>

<sup>1</sup>Estudante do Curso de Economia – ICS- UFOPA; E-mail: andrezzaalves68@gmail. <sup>2</sup> MSc. em Estatística – ICS - UFOPA.  
E-mail: tarcisiolobato@yahoo.com.br.

**RESUMO:** Devido a existência várias técnicas estatísticas para serem aplicadas em determinado modelo, o presente resumo tem como objetivo descrever o passo a passo para realização de uma técnica do método de análise fatorial no SPSS utilizando com base de dados variáveis referente aos países do Continente Americano, de modo que os dados foram obtidos no portal do IBGE. Uma vez que para realização do estudo foi realizado pesquisas bibliográficas, usando a técnica de análise multivariada com enfoque na análise fatorial usando o programa SPSS.

**Palavras-chave:** análise multivariada; análise fatorial; SPSS.

## INTRODUÇÃO

Há diversas técnicas estatísticas para serem aplicadas, e dentre elas temos a Análise Multivariada, que consiste em diversos testes. De forma que a estatística se tornou um instrumento essencial para a análise e avaliação de dados, em diversas áreas do conhecimento.

Para Corrar et al (2007) a Análise Multivariada refere-se a um conjunto de métodos estatísticos que torna possível a análise simultânea de medidas múltiplas para cada indivíduo, objeto ou fenômeno observado. Nesse sentido, os métodos que permitem a análise simultânea de mais de duas variáveis podem ser considerados como integrantes da análise multivariada. Já que a análise multivariada corresponde às diversas abordagens analíticas que consideram o comportamento de muitas variáveis simultaneamente, torna-se viável sintetizar as diferenças básicas das análises univariada (quando possui apenas uma variável), bivariada (duas variáveis) e multivariada (mais de duas variáveis).

Segundo Vicini (2005) análise fatorial é constituída por um conjunto de técnicas estatísticas, e tem como finalidade diminuir o número de variáveis iniciais com o menor prejuízo possível de informação. Isto é, a análise fatorial procura identificar os fatores num conjunto de medidas realizadas. Notando que esses fatores identificados pela análise fatorial são uma descoberta feita pelo pesquisador. Apresentando várias medidas e não sendo possível acomodar quais variáveis poderão ser concentradas num fator, pois é a análise fatorial que proporciona isso, uma vez que ela admite identificar novas variáveis em um número abreviado em relação às variáveis iniciais, sem uma perda significativa de informação contida nos dados originais.

A análise fatorial pode ser usada em diversos exemplos do cotidiano, de modo que ela reduz a amostra e identifica quais variáveis influenciam com mais potência no modelo. A partir deste pressuposto, busca-se descrever o modelo de análise fatorial e como realiza- lá no software SPSS, ressaltando o passo a passo para realização da mesma e explicando os resultados expostos. Assim como verificar quais variáveis afetam com mais intensidades os países do Continente Americano.

A principal função das diferentes técnicas de análise fatorial é reduzir uma grande quantidade de variáveis observadas em um número menor de fatores. Um fator é uma combinação linear das variáveis (estatísticas) originais. Sendo assim, a análise fatorial aborda o problema de analisar a estrutura das inter-relações (correlações) entre um grande número de variáveis, definindo um conjunto de dimensões latentes comuns, chamados fatores (HAIR et al, 2005).

## PROCEDIMENTOS METODOLÓGICOS

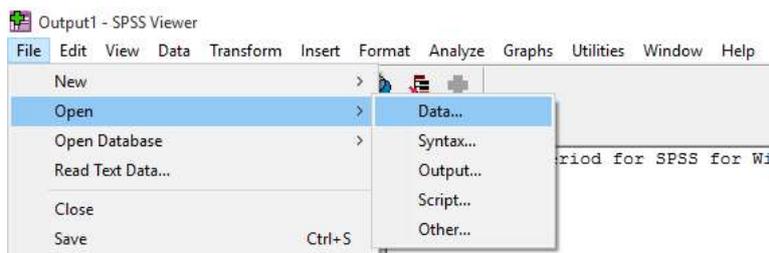
Este trabalho foi feito por meio de pesquisas bibliográficas, usando a técnica de análise multivariada com enfoque na análise fatorial usando o programa SPSS, no qual é um software aplicativo do tipo científico que tinha originalmente o nome era acrônimo de *Statistical Package for the Social*, que corresponde um pacote estatístico para as ciências sociais que inclui: aplicação analítica, *Data Mining*, *Text Mining* e estatística.

Esta pesquisa é referente ao continente americano, no qual possui uma área de 42 189 120 km<sup>2</sup>, com uma população de 902 892 047 habitantes possuindo uma densidade demográfica de 21 hab./km<sup>2</sup> sendo que a região possui um número de 35 países: Antígua e Barbuda, Argentina, Bahamas, Barbados, Belize, Bolívia, Brasil, Canadá, Chile, Colômbia, Costa Rica, Cuba, Dominica, El Salvador, Equador, Estados Unidos, Granada, Guatemala, Guiana, Haiti, Honduras, Jamaica, México, Nicarágua, Panamá, Paraguai, Peru, República Dominicana, Santa Lúcia, São Cristóvão e Névis, São Vicente e Granadinas, Suriname, Trinidad e Tobago, Uruguai, Venezuela (IBGE, 2017).

Foram utilizados os dados de Gastos públicos com educação, Gastos públicos com saúde, PIB per capita, Total da exportação, Total da importação, Índice de desenvolvimento humano, Esperança de vida ao nascer, Áreas cultivadas, Áreas protegidas no total do território nacional, Taxa bruta de mortalidade, Taxa bruta de natalidade, Taxa média anual do crescimento da população, Número de computadores pessoais, Usuários com acesso à internet, coletados no portal do Instituto Brasileiro de Geografia Estatística (IBGE, 2017).

## RESULTADOS E DISCUSSÃO

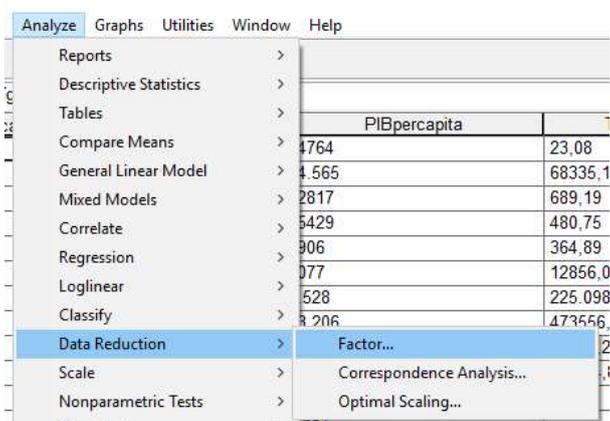
Para Realizar o método de análise fatorial no SPSS, devemos: Abrir o programa, ir em *Open; Data* e escolher o arquivo que deseja usar.



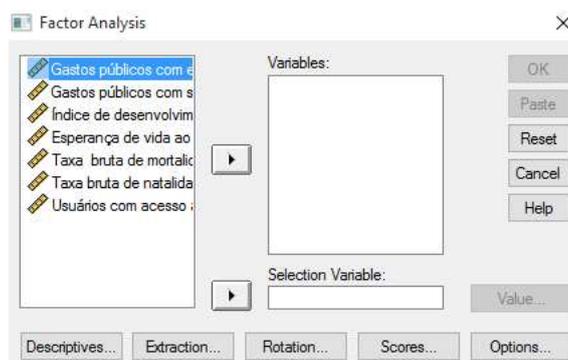
Depois de abrir o arquivo, os dados apareceram na tela inicial do SPSS:

	PaisesdaAmerica	Gastospúbl	Gastospúbl	PIBpercapita
1	Antigua e Barbuda	2,5	4,9	14764
2	Argentina	5,8	7,3	14.565
3	Bahamas	.	7,3	22817

Após a visualização da tabela, vamos realizar a análise fatorial, para isso: aperte em *Analyse, Data Reduction, Factor*.



Depois basta escolher as variáveis que deseja na sua análise e em seguida escolha as opções de suas análises e aperte *ok*.



Ressaltando que na parte de *descriptives*, podem-se escolher as formas descritivas das variáveis e sua matriz de correlação, na parte de extração, poderá escolher o método utilizado para extração dos dados, na parte de rotação terá a possibilidade de escolher qual o método utilizado para rotacionar o modelo, ressaltando que o mais usado é o *Varimax*. Já na opção de *Scores*, poderá escolher em qual método deseja salvar os resultados e por fim na parte de *Options* poderá excluir os casos que têm valores perdidos para qualquer das variáveis usadas em qualquer das análises, substituir os valores perdidos com a variável média e entre outras preferências.

Os resultados obtidos mostram média das variáveis e o desvio padrão.

	Mean	Std. Deviation
Gastos públicos com educação	5,004	2,1695
Gastos públicos com saúde	7,437	2,5963
Índice de desenvolvimento humano	,74556	,073804
Esperança de vida ao nascer	75,144	3,6135
Taxa bruta de mortalidade	6,37	1,363
Taxa bruta de natalidade	18,52	4,726
Usuários com acesso à internet	52,127	16,2501

Quadro 1. Análise Descritiva dos dados.

De acordo Hair et. al (1998) os Testes *Kaiser-Meyer-Olkin* (KMO) e de Esfericidade de Bartlett, sugerem qual é o grau de suscetibilidade ou o ajuste dos dados à análise fatorial, ou seja, mostra o nível de confiança que se pode confiar dos dados. Uma vez que, o KMO exibe valores normalizados (entre 0 e 1,0) e mostra qual é a proporção da variância que as variáveis proporcionam em comum ou a proporção desta que são devidas a fatores comuns.

Para Pereira (2001) elucida que valores próximos de 1,0 sugerem que o método de análise fatorial é perfeitamente adequado para o tratamento dos dados. Mas valores menores que 0,5, sugerem a inadequação do método.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,685
Bartlett's Test of Sphericity	Approx. Chi-Square	112,307
	Df	21
	Sig.	,000

Quadro 2. Valor KMO e teste de esfericidade de Bartlett

Desta forma, o valor do modelo deu 0,685, o que indica uma razoável adequação de possibilidades de tratamento dos dados com o método citado.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,808	54,396	54,396	3,808	54,396	54,396	3,296	47,079	47,079
2	1,180	16,853	71,248	1,180	16,853	71,248	1,692	24,170	71,248

3	,814	11,634	82,883
4	,658	9,403	92,286
5	,324	4,633	96,919
6	,155	2,221	99,140
7	,060	,860	100,000

Quadro 3: Valores próprios e percentagens de variância explicada das 6 componentes principais amostrais.

Analisando quadro 3, observou-se que a variável Gastos públicos com educação explica 47,079 % do modelo e que junto com a Gastos públicos com saúde explicam 71,248 do modelo. Ressaltando que essas duas variáveis são as que mais explicam o modelo analisado.

	Initial	Extraction
Gastos públicos com educação	1,000	,776
Gastos públicos com saúde	1,000	,438
Índice de desenvolvimento humano	1,000	,932
Esperança de vida ao nascer	1,000	,706
Taxa bruta de mortalidade	1,000	,527
Taxa bruta de natalidade	1,000	,826
Usuários com acesso à internet	1,000	,782

Quadro 4: Comunalidades das variáveis

Schawb (2007) elucida que as *Communalities* concebem a proporção da variância para cada variável compreendida na análise que é explicada pelos componentes extraídos. A partir disso, observou-se que as variáveis que apresentaram uma *communalities* baixa foram Gastos públicos com saúde e Taxa bruta de mortalidade e as demais apresentaram uma *communalities* alta. Lembrando que disso, baixa *communalities* entre um grupo de variáveis é um sinal de que elas não estão linearmente correlacionadas o que indica que elas não devem ser contidas na análise fatorial.

	Component	
	1	2
Gastos públicos com educação	-,046	,880
Gastos públicos com saúde	,518	,411
Índice de desenvolvimento humano	,947	,188
Esperança de vida ao nascer	,827	,148
Taxa bruta de mortalidade	,296	,663
Taxa bruta de natalidade	-,758	-,502
Usuários com acesso à internet	,884	-,011

Quadro 5- Matriz de pesos fatoriais rotacionada segundo critério Varimax.

Analisando o quadro 5, e lembrando que os pesos fatoriais mais elevados em valor absoluto identificam o fator a que cada variável se associa, observou-se que o fator 1 é formado pelas variáveis Gastos públicos com saúde, Índice de desenvolvimento humano, Esperança de vida ao nascer, Taxa bruta de natalidade e Usuários com acesso à internet, ressaltando que 4 variáveis são correlacionadas positivamente e 1 correlacionadas negativamente. O fator 2 é composto pelas variáveis Gastos públicos com educação e Taxa bruta de mortalidade, sendo que ambas são correlacionadas positivamente.

## CONCLUSÕES

A partir dos resultados, conclui-se que as variáveis Gastos públicos com educação e Taxa bruta de mortalidade poderiam ser excluídos do modelo, pois poderia que o mesmo tendesse a uma explicação melhor. Dos dados apresentados, observou-se que depois da realização da análise fatorial, o modelo poderia ser dividido em dois fatores.

Ressaltando que no continente americano as variáveis que tendem a ter mais impacto são gastos públicos com educação e Gastos públicos com saúde, pois as mesmas explicam cerca de 71,5 do modelo. Outro ponto a se

destacar é seria o teste KMO, que mostrou que o modelo é m modelo razoável. Sendo assim, par realização e um próximo estudo, poderia levar em consideração outras variáveis ou até mesmo a exclusão de algumas variáveis para que o modelo se torne mais explicado e com menos erros.

#### REFERÊNCIAS

CORRAR, L.J.; PAULO, E.; DIAS FILHO, J.M. **Análise multivariada: para os cursos de Administração, Ciências Contábeis e Economia**. São Paulo: Atlas, 2007.

HAIR, J. F., Jr.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. **Análise multivariada de dados**. 5. ed. Porto Alegre, 2005.

IBGE. **Países**. 2017. Disponível em <<https://paises.ibge.gov.br/>>. Acesso em: 14 de setembro de 2017.

VICINI, Lorena. **Análise multivariada da teoria à prática**. RS, 2005.