

Análise Fatorial

Bruna e Denis

01/10/2021

Baseline

- Tipo de análise;
- Quando é utilizada;
- Para que tipos de dados devem ser aplicados;
- Passo a passo dessa análise de forma resumida;
- Exemplo (artigo).

O que é análise fatorial?

Técnica de **interdependência** cujo propósito principal é definir a estrutura inerente entre as variáveis de análise.

Pode ser utilizada para analisar inter-relações entre um grande número de variáveis e explicar essas variáveis em termos de suas dimensões inerentes comuns (fatores).

Resumir ou reduzir a quantidade de variáveis

Criar escalas múltiplas ou escores

A análise fatorial apresenta diversas maneiras de representação desses grupos de variáveis para uso em outras técnicas multivariadas.

PARTE 1: MATRIZ ORIGINAL DE CORRELAÇÃO

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
V_1 Nível de preço	1,000								
V_2 Pessoal da loja	0,427	1,000							
V_3 Política de devolução	0,302	0,771	1,000						
V_4 Disponibilidade do produto	0,470	0,497	0,427	1,000					
V_5 Qualidade do produto	0,765	0,406	0,307	0,427	1,000				
V_6 Profundidade de diversidade	0,281	0,445	0,423	0,713	0,325	1,000			
V_7 Amplidão da diversidade	0,345	0,490	0,471	0,719	0,378	0,724	1,000		
V_8 Serviço interno	0,242	0,719	0,733	0,428	0,240	0,311	0,435	1,000	
V_9 Atmosfera da loja	0,372	0,737	0,774	0,479	0,326	0,429	0,466	0,710	1,000

PARTE 2: MATRIZ DE CORRELAÇÃO DE VARIÁVEIS APÓS AGRUPAMENTO DE ACORDO COM ANÁLISE FATORIAL

	V_3	V_8	V_9	V_2	V_6	V_7	V_4	V_1	V_5
V_3 Política de retorno	1,000								
V_8 Serviço interno	0,773	1,000							
V_9 Atmosfera da loja	0,771	0,710	1,000						
V_2 Pessoal da loja	0,771	0,719	0,737	1,000					
V_6 Profundidade de diversidade	0,423	0,311	0,429	0,445	1,000				
V_7 Amplidão de diversidade	0,471	0,435	0,466	0,490	0,724	1,000			
V_4 Disponibilidade do produto	0,427	0,428	0,479	0,497	0,713	0,729	1,000		
V_1 Nível de preço	0,302	0,242	0,372	0,427	0,281	0,354	0,470	1,000	
V_5 Qualidade do produto	0,307	0,240	0,326	0,406	0,325	0,378	0,427	0,765	1,000

Nota: Áreas sombreadas representam variáveis agrupadas por análise fatorial.

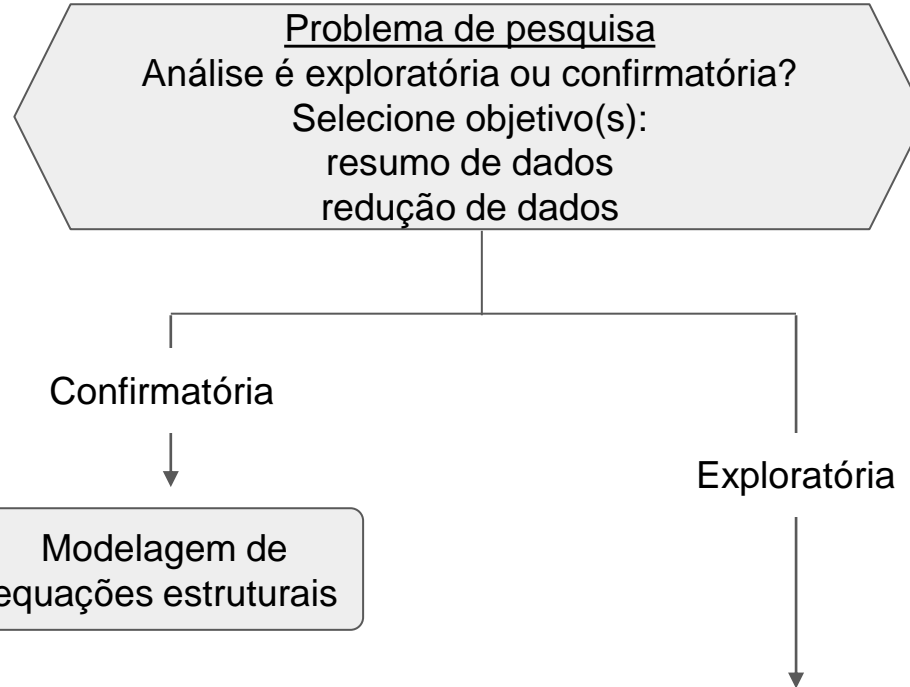
FIGURA 3-1 Exemplo ilustrativo do uso de análise fatorial para identificar estrutura dentro de um grupo de variáveis.

Passo a passo

Hair (2009) propõe sete estágios da análise fatorial:

1. Objetivos da análise
2. Planejamento
3. Suposições
4. Determinação de fatores e avaliação do ajuste geral
5. Interpretação dos fatores
6. Validação
7. Usos adicionais dos resultados

Estágio 1



“ideias preconcebidas sobre
a real estrutura dos dados”

“testar hipóteses”

Confirmatória

Modelagem de
equações estruturais

Exploratória

“útil na busca da estrutura
em um conjunto de
variáveis ou como método
de redução de dados”

resumo de dados - “variáveis individuais são agrupadas e então vistas não por aquilo que elas representam individualmente, mas por aquilo que representam coletivamente na expressão de um conceito.”

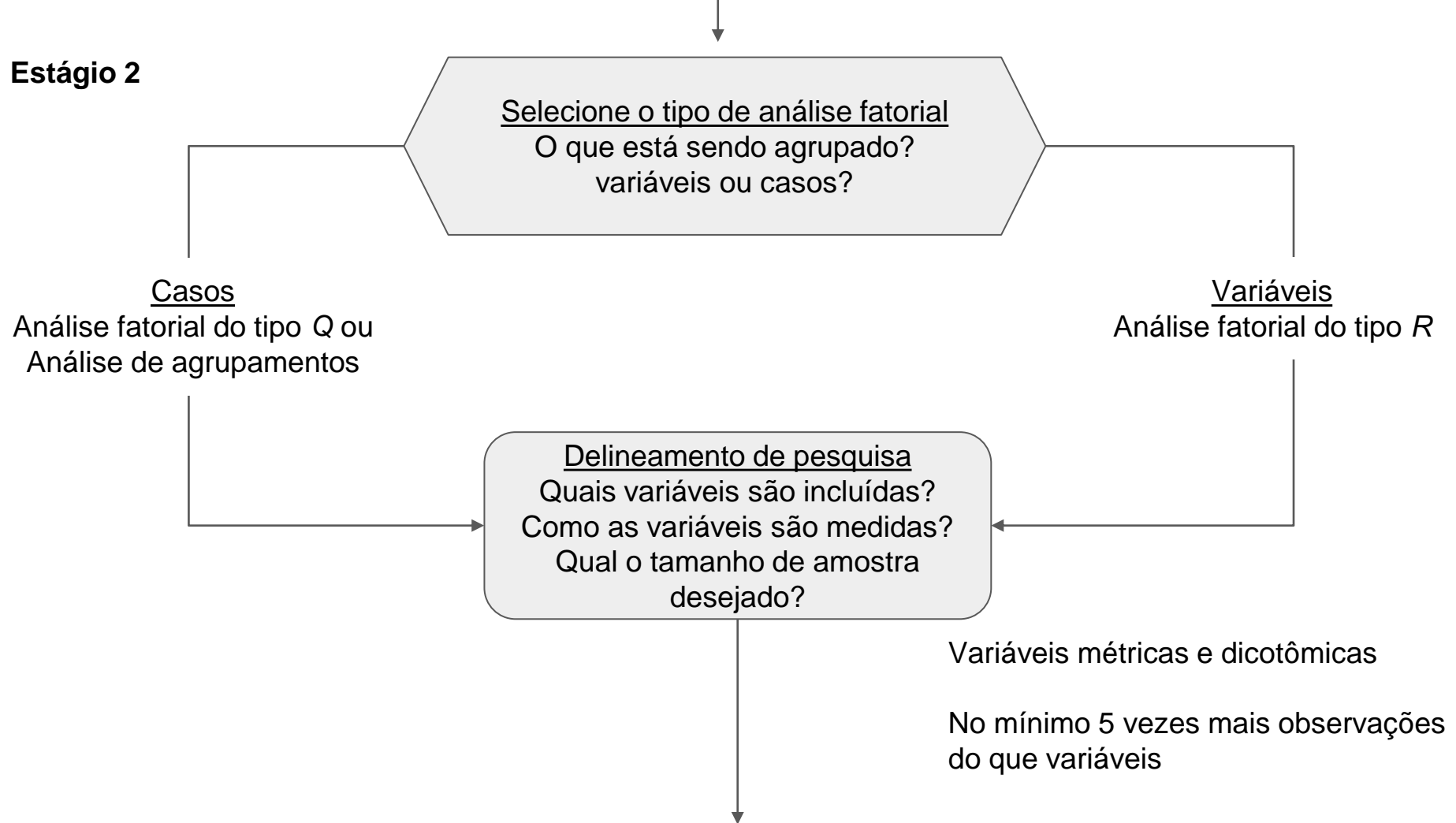
estrutura - “relações entre variáveis, viabilizando a especificação de um número menor de dimensões (fatores) representando o conjunto original de variáveis.”

redução de dados

(1) identificação de variáveis representativas a partir de um conjunto maior de variáveis para uso em análises multivariadas subsequentes

(2) criação de um conjunto inteiramente novo de variáveis, muito menor, para substituir parcial ou completamente o conjunto original de variáveis

Estágio 2



<i>Respondente</i>	<i>Variáveis</i>		
	V_1	V_2	V_3
A	7	7	8
B	8	6	6
C	2	2	3
D	3	1	1

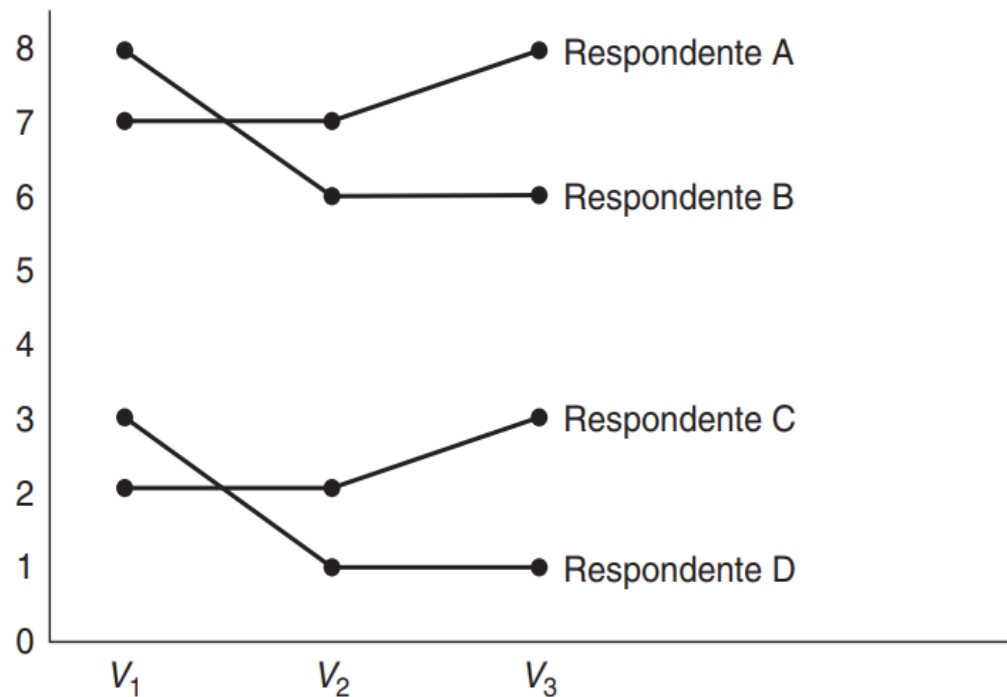


FIGURA 3-3 Comparações de perfis de escore para análise fatorial do tipo *Q* e análise de agrupamentos.

Estágio 3

Suposições

Considerações estatísticas de normalidade,
linearidade e homoscedasticidade
Homogeneidade da amostra
Conexões conceituais

- Uma forte fundamentação conceitual é necessária para embasar a suposição de que existe uma estrutura anterior à análise fatorial (*questões conceituais*);
- Um pouco de multicolinearidade é desejável, pois o objetivo é identificar conjuntos de variáveis inter relacionadas e é necessário que sejam suficientemente correlacionadas umas com as outras para produzir fatores representativos;
- Medidas de intercorrelação:
 - correlações parciais;
 - Teste de Esfericidade de Bartlett: teste para correlações significantes entre pelo menos algumas das variáveis (sign. < 0,05);
 - Medida de adequação da amostra (MSA): quantifica o grau de intercorrelações entre as variáveis e a adequação da análise fatorial (geral e específica)

Estágio 3

Estágio 4

Seleção de um método fatorial
A variância total é analisada, ou apenas a variância comum?

Variância total

Extraia fatores com
análise de componentes

- considera a variância total;
- redução de dados é uma preocupação prioritária;
- conhecimento prévio sugere que variância específica e de erro representam uma proporção relativamente pequena da variância total;
- deriva fatores que contêm pequenas proporções de variância única e, em alguns casos, variância de erro;

Variância comum

Extraia fatores com
análise de fatores comuns

- considera apenas variância em comum ou compartilhada;
- o objetivo prioritário é identificar as dimensões ou construtos latentes;
- pouco conhecimento sobre a quantia de
- variância específica e de erro e, logo, não são de interesse na definição da estrutura das variáveis;

Especificação da matriz fatorial
Determine o número de fatores
a serem mantidos

- raiz latente (autovalor >1)
- *priori*
- porcentagem da variância (60% +)
- teste *scree* (gráfico de cotovelo)
- heterogeneidade das observações

Estágio 5

Seleção de um método rotacional
Os fatores devem ser correlacionados (obliquos)
ou não correlacionados (ortogonais)?

Métodos ortogonais

varimax
equimax
quartmax

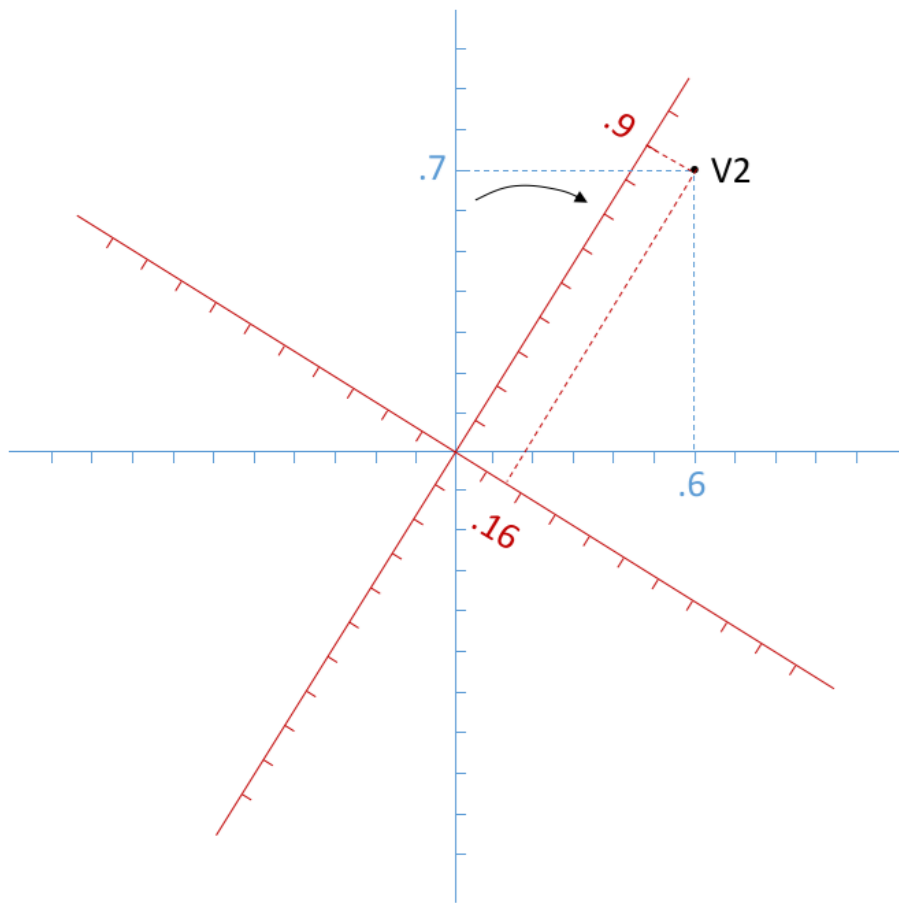
varimax: mais utilizado. rotaciona o fator inicial de modo que uma variável tenha carga alta em um fator e cargas tão baixas quanto possível em todos os outros fatores

Rotação de fatores: os eixos de referência dos fatores são rotacionados em torno da origem até que alguma outra posição seja alcançada;

Métodos oblíquos

oblimin
promax
orthoblique

Variáveis	Cargas fatoriais não rotacionadas		Cargas fatoriais rotacionadas	
	Fator I	Fator II	Fator I	Fator II
V1	.50	.80	.03	.94
V2	.60	.70	.16	.90
V3	.90	-.25	.95	.24
V4	.80	-.30	.84	.15
V5	.60	-.50	.76	-.13



Seleção de um
método fatorial

Não

Interpretação da matriz fatorial rotacionada

Podem ser encontradas cargas significantes?

Os fatores podem ser nomeados?

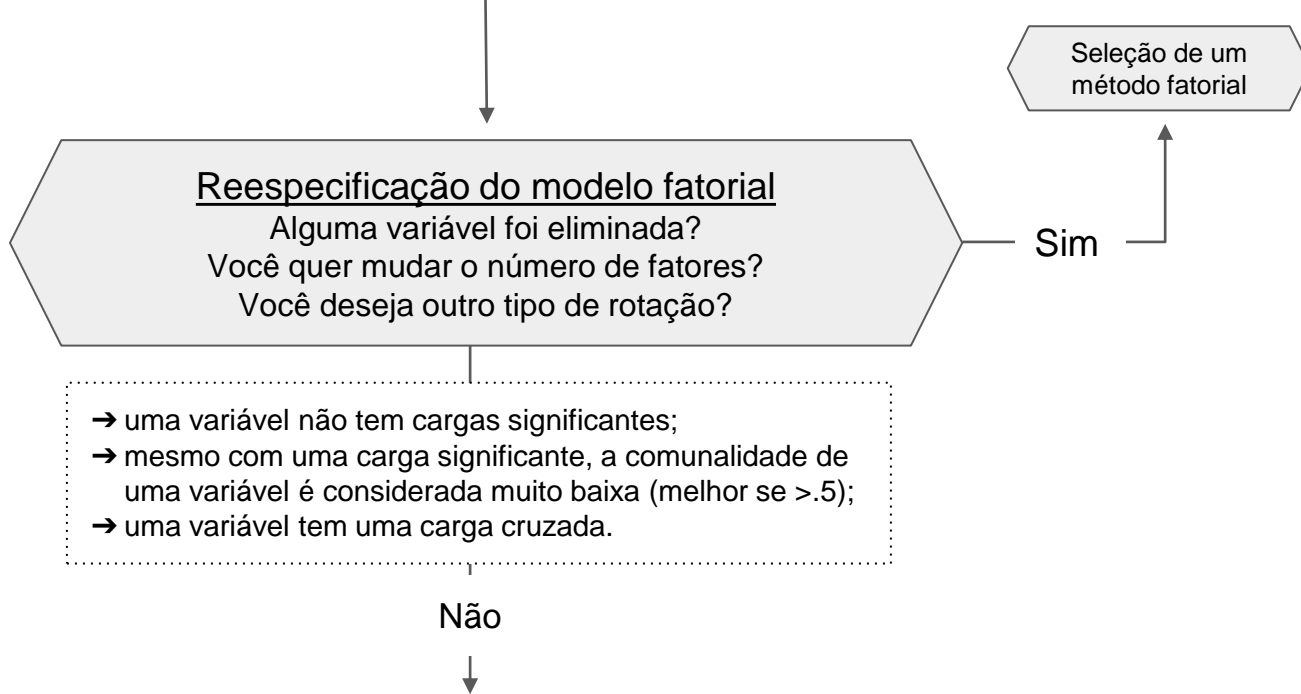
As comunalidades são suficientes?

- Os fatores são dispostos como colunas representando as cargas de um único fator.
- A avaliação dos fatores deve começar com a primeira variável no primeiro fator e se mover horizontalmente da esquerda para a direita, procurando a carga mais alta para aquela variável em qualquer fator, identificando-a. Variáveis com mais de uma carga significativa é chamada carga cruzada.
- A comunalidade de uma variável é a estimativa de sua variância compartilhada, ou em comum, entre as variáveis como representadas pelos fatores obtidos ($< .5$ insuficiente)

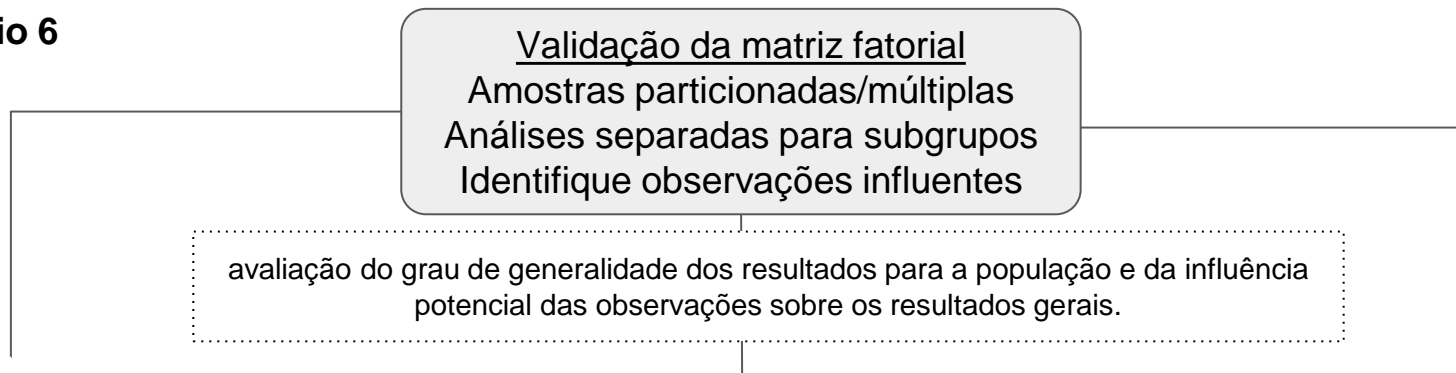
Sim

Estágio 5

Estágio 5



Estágio 6



Estágio 7

- Identificar combinações lógicas de variáveis e entender melhor as inter relações entre variáveis (resumo dos dados);
- Identificar variáveis apropriadas para a aplicação subsequente em outras técnicas estatísticas (redução de dados);

Seleção de variáveis de substituição

- Selecionar a variável com a maior carga fatorial como uma representativa substituta para uma dimensão fatorial particular;
- Decisão pode ser baseada *a priori*.

Computação de escores fatoriais

- medida composta de cada fator computadas para cada indivíduo;
- o escore fatorial representa o grau em que cada indivíduo tem escore elevado no grupo de itens que têm cargas elevadas em um fator;
- é computado com base nas cargas fatoriais de todas as variáveis no fator;


Criação de escalas múltiplas

- combinação de diversas variáveis individuais em uma única medida composta;
- todas as variáveis com cargas elevadas em um fator são combinadas, e o escore médio é usado como uma variável de substituição.
- Supera o erro de medida;
- Representa múltiplos aspectos de um conceito com uma medida única;
- é calculada combinando-se apenas variáveis selecionadas.



Article

A Perspective on Inhabited Urban Space: Land Use and Occupation, Heat Islands, and Precarious Urbanization as Determinants of Territorial Receptivity to Dengue in the City of Rio De Janeiro

Jefferson Pereira Caldas Santos ^{1,*} , Nildimar Alves Honório ^{2,3}, Christovam Barcellos ⁴
and Aline Araújo Nobre ⁵

¹ Centro de Inovação em Biodiversidade e Saúde, Instituto de Tecnologia em Fármacos, Fundação Oswaldo Cruz, Rio de Janeiro 22775-903, Brazil

² Laboratório de Mosquitos Transmissores de Hematozoários, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro 21040-900, Brazil; nildimar.honorio@ioc.fiocruz.br

³ Núcleo Operacional Sentinela de Mosquitos Vetores-Nosmove/Fiocruz, Fundação Oswaldo Cruz, Rio de Janeiro 21040-900, Brazil

⁴ Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro 21040-900, Brazil; xris@icict.fiocruz.br

⁵ Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro 21040-900, Brazil; aline.nobre@fiocruz.br

* Correspondence: jefferson.santos@far.fiocruz.br

In this context, this study aimed to stratify the city of Rio de Janeiro into areas of dengue receptivity through innovative “territorial indicators” whose construction takes into account the actual reality of the occupation of the territory.

Table 1. Territorial indicators, and their descriptive statistics and data sources.

Code	Indicator	Construction	Unit	Min–Max	Mean (SD)	Period	Source
Ind 01	Percentage of occupied area	Occupied Area/Total Area	%	3.38–100	76.013 (24.524)	2014	Municipality
Ind 02	Percentage of residential area	Residential Area/Total Area	%	0.74–100	64.878 (24.117)	2014	Municipality
Ind 03	Net demographic density	Total population/Residential area	inhab/km ²	1372.74–81,171.23	16,778.461 (10,518.778)	2014	Municipality
Ind 04	Percentage of area of substandard clusters	Substandard Cluster Area/Residential Area	%	0–98.38	11.187 (14.7)	2014	Municipality
Ind 05	Percentage of are with strategic points	Strategic Points Area/Occupied Area	%	0–14.81	1.154 (2.547)	2014	Municipality
Ind 06	Mean verticalization	Total building heights/Total buildings	m	2.45–36.69	6.37 (4.778)	2014	Municipality
Ind 07	Percentage of vegetation	Vegetation Area/Total Area	%	0–94.34	25.108 (24.029)	2014	Municipality
Ind 08	Number of neighboring districts	Number of neighbors through neighborhood matrix	districts	1–13	4.375 (1.955)	2014	Own
Ind 09	Border perimeter with neighboring districts	Total length of boundary line	m	12.33–32048.3	7095.125 (4658.223)	2014	Own
Ind 10	Mean daytime surface temperature	Mean daytime temperature in the occupied area	°C	23.69–33.48	30.22 (2.623)	2008–2014	MODIS/IRI
Ind 11	Mean nighttime surface temperature	Mean nighttime temperature in the occupied area	°C	19.24–22.13	21.244 (0.733)	2008–2014	MODIS/IRI
Ind 12	Monthly cumulative rainfall	Mean cumulative rainfall in occupied area	mm ³	85.13–153.31	100.058 (10.844)	2008–2014	Municipality
Ind 13	Percentage of households with unpaved streets	(v015 + v017 + v019/v001)—Spreadsheet surrounding01	%	0–63.63	4.831 (10.452)	2010	2010 Census
Ind 14	Percentage of households with no tree-lined streets	(v045 + v047 + v049/v001)—Spreadsheet surrounding01	%	0–77.28	19.869 (17.862)	2010	2010 Census
Ind 15	Percentage of households with streets without manholes	(v033 + v035 + v037/v001)—Spreadsheet surrounding01	%	0–91.96	12.342 (15.126)	2010	2010 Census
Ind 16	Percentage of households with streets with exposed trash	(v056 + v058 + v060/v001)—Spreadsheet surrounding01	%	0–40.34	4.038 (5.584)	2010	2010 Census
Ind 17	Percentage of households with streets with open sewage	(v050 + v052 + v054/v001)—Spreadsheet surrounding01	%	0–31.44	4.123 (5.534)	2010	2010 Census 2010

Ind: Indicator; Inhab/km²: Inhabitant per square kilometer.

2.4. Data Analysis

We created dengue receptivity dimensions through principal component analysis (PCA) using the 17 indicators described in Table 1. PCA is a multivariate analysis technique that aims to transform the original, possibly correlated variables into components that are linear orthogonal combinations of these variables, to reduce the dimensions with the least possible loss of information [40]. The indicators were standardized, and we used the Kaiser criterion to identify the components to be selected, keeping only those with eigenvalues >1.0 . The importance of each major component is assessed by the proportion of the total variance explained by the component. The load of each indicator was used to determine its importance in the component construction. In order to aid interpretation, components were named based on the indicators with the highest input.

We constructed the dengue receptivity index using multiple criteria analysis. This procedure involves map algebra, in which different information sheets are cross-referenced with their weights and scores, resulting in the synthesis map. Criteria must be standardized to carry out this integration, standardizing the units of all maps [41]. In this study, we used the components resulting from the principal component analysis as the information sheets to generate the dengue receptivity map. This map was generated by a multiple criteria analysis using a weighted linear combination method, in which each component was normalized and weighted according to its correlation index with the mean rate for the entire study period (2008 to 2018). Subsequently, the receptivity index was divided into quintiles, and the resulting map was constructed considering five classes (very high, high, medium, low, and very low).

The dengue incidence rate and *Aedes* egg density values were interpolated by the weighted inverse distance weighting (IDW) method to generate a smooth and continuous layer from which isolines of equal value were extracted. The vector-format *Aedes* interpolated (smoothed) incidence rate and egg

Table 2. Factor loadings, eigenvalue, and explained variance of principal component analysis.

Indicators	Factor Loadings			
	Comp 1	Comp 2	Comp 3	Comp 4
Ind 01	0.91 *	0.03	-0.25	0.05
Ind 02	0.86 *	0.05	-0.18	0.06
Ind 03	0.26	-0.67 *	-0.03	0.52 *
Ind 04	0.05	-0.28	0.03	0.78 *
Ind 05	0.14	0.1	0.35	0.02
Ind 06	0	-0.67 *	0	-0.01
Ind 07	-0.88 *	0.02	0.25	-0.03
Ind 08	0.45	0.14	0.64 *	-0.16
Ind 09	0.28	0.41	0.50 *	-0.14
Ind 10	0.78 *	0.44	0.06	0.08
Ind 11	0.87 *	0.19	-0.03	0.05
Ind 12	-0.52	-0.52	0.31	0.03
Ind 13	-0.58	0.48	-0.34	0.09
Ind 14	-0.08	0.59 *	-0.22	0.25
Ind 15	-0.57	0.58 *	-0.33	0.1
Ind 16	-0.09	0.36	0.44 *	0.43 *
Ind 17	-0.29	0.52	0.28	0.39 *
Eigenvalue	5.11	2.99	1.6	1.36
% of variance	30.06	17.6	9.44	7.99

Ind: Indicator; Comp: Components; * Indicators with the greatest contribution in each component.

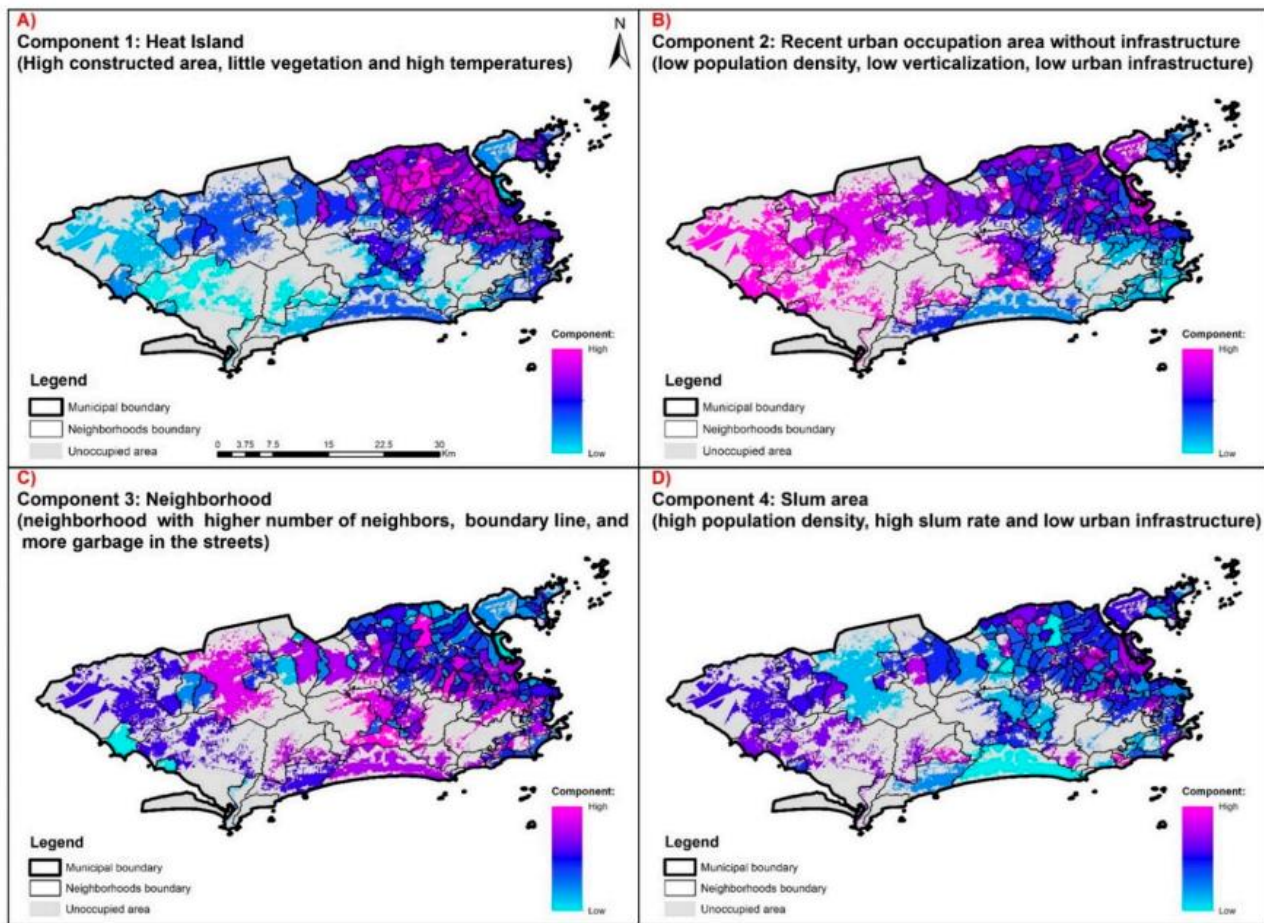


Figure 4. Spatialization of components from principal component analysis in Rio de Janeiro city: (A) component 1, (B) component 2, (C) component 3, and (D) component 4.

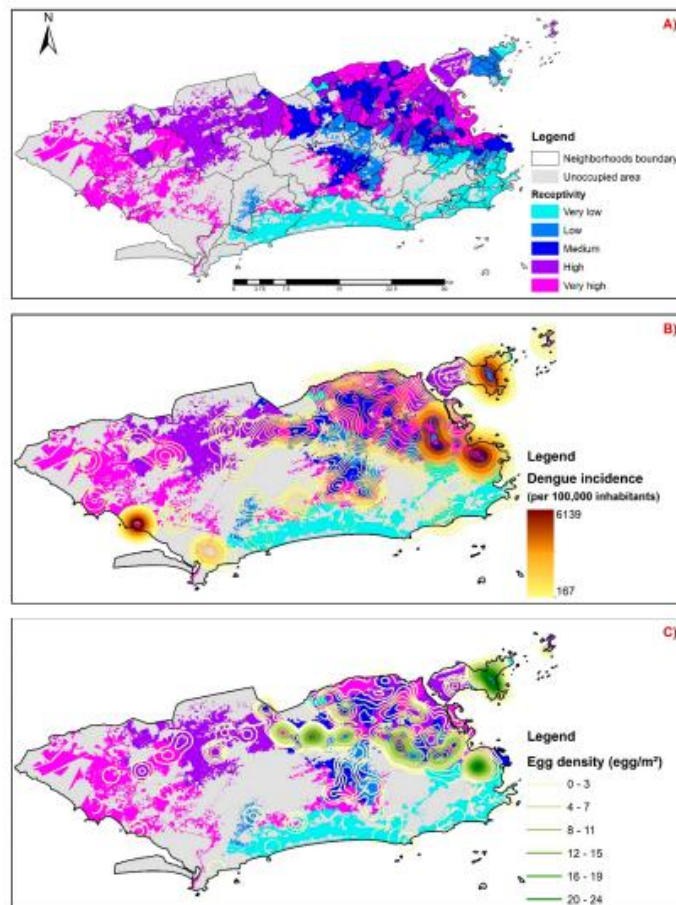


Figure 5. Map of dengue receptivity of Rio de Janeiro versus mean dengue incidence rate (2008 to 2018) and *Aedes* egg density index (2013–2014); (A) Mapping the receptivity index by neighborhoods in the city of Rio de Janeiro; (B) Mapping the density of the average incidence of dengue overlapping the receptivity index by neighborhoods in the city of Rio de Janeiro; (C) Mapping the density of *Aedes* eggs overlaid the receptivity index by neighborhoods in the city of Rio de Janeiro.

Table 3. Spearman correlation coefficient between receptivity index, dengue incidence rate, and *Aedes* egg density index (EDI) in the neighborhoods of Rio de Janeiro City.

	Incidence											EDI	
Period	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2008–2018	2013–2014
Serotype	Denv 2	Denv 2	Denv 2	Denv 1	Denv 4	Denv 4	Denv 4	Denv 1	Denv 1	Denv 2	Denv 2		
Component 1	0.11	0.02	−0.15 *	−0.19 *	0.31 *	−0.13	0.03	0.27 *	0.20 *	0.12	0.12	0.12	0.07
Component 2	0.29 *	0.08	−0.21 *	0.24 *	0.41 *	−0.16 *	0.08	0.17 *	0.46 *	0.22 *	0.36 *	0.37 *	0.28 *
Component 3	0.04	−0.01	−0.01	−0.02	−0.02	−0.04	0.02	0.03	0.01	−0.16 *	−0.01	−0.05	−0.05
Component 4	0.14 *	0.02	−0.09	0.00	0.01	−0.12	−0.05	0.06	0.19 *	0.17 *	0.19 *	0.09	−0.06
Receptivity	0.28 *	0.08	−0.24 *	0.14	0.44 *	−0.22 *	0.05	0.23 *	0.47 *	0.30 *	0.39 *	0.35	0.21 *

* Significant with a significance level of 5%.

Obrigada(o)!