

Aula 4: Descrever a Diversidade na Realidade Social

Análise de Clusters e Análise Factorial

Docente: Amílcar Moreira

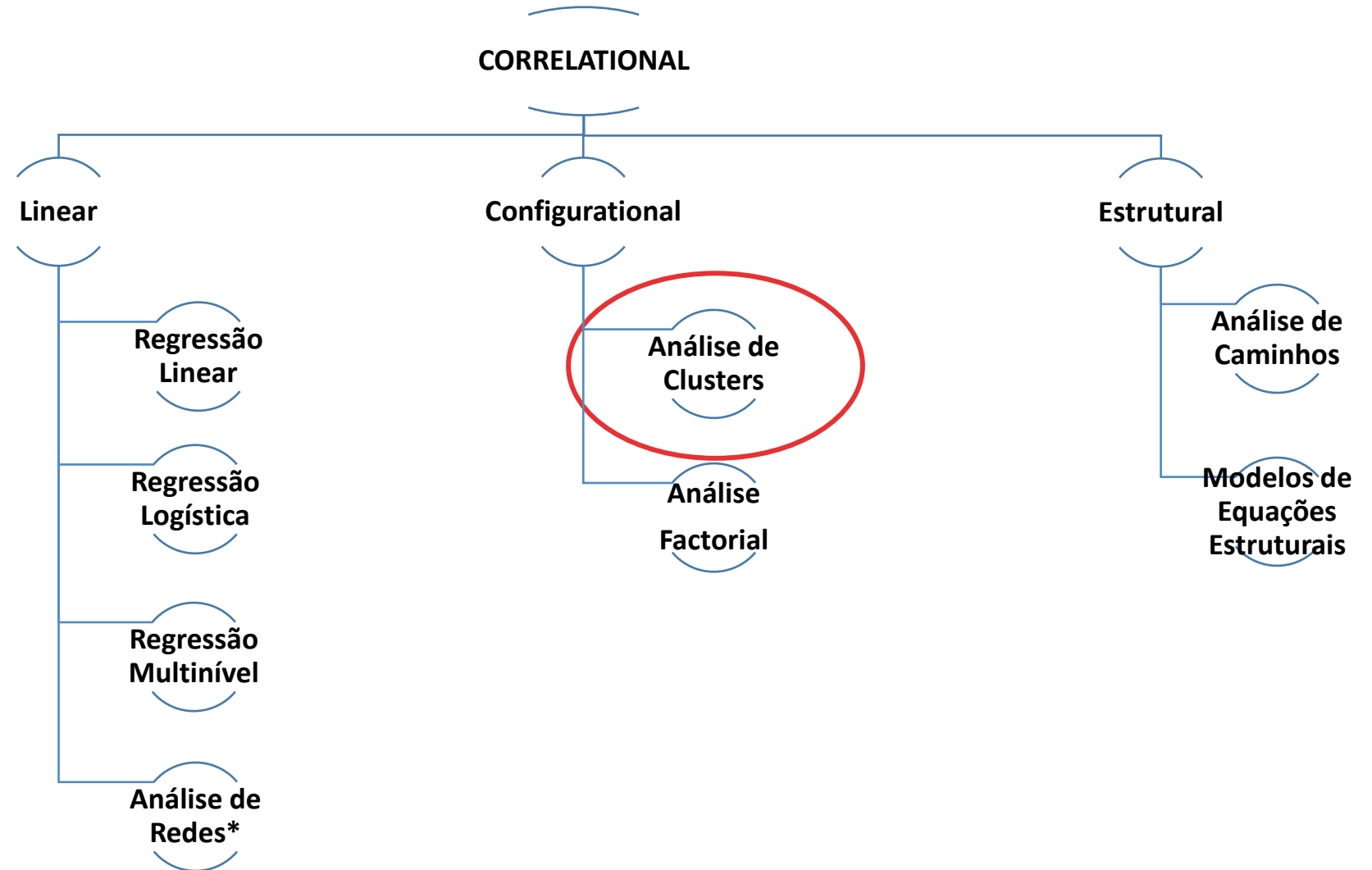
Dia & Hora: 02/10/2023, 20:30-22:30

Sala: Q6, IAPMEI Room

Métodos e Técnicas de Investigação

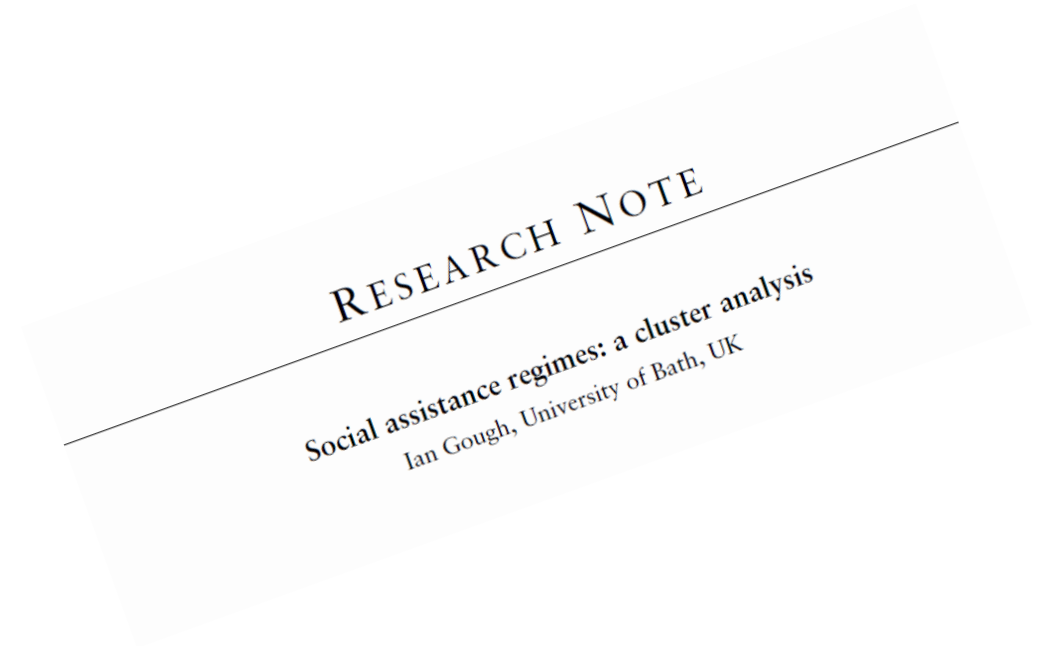
Aula 4: Descrever a Diversidade na Realidade Social

- As you can see, there is large variety of Correlational Research Designs...



Análise de Clusters

- **Origem na Antropologia**
- **Especialmente útil para identificar a existência de grupos de observações (indivíduos/empresas/países) que partilham entre si um conjunto de características.**



Um exemplo prático

- **Imaginemos que queremos identificar diferentes modelos de proteção social num conjunto de países.**

Um exemplo prático:

- Imaginemos que queremos identificar diferentes modelos de proteção social num conjunto de países
- Para o fazer, recolhemos dados que nos permitam tipificar diferenças nos modelos de proteção social.

Table 1

Definition of 15 indicators for developmental welfare states

Indicator	Definition	Principal data resource
Governmental social expenditure	Percentage of 'social security and welfare' expenditure in total public expenditure	IMF
Social investment	Percentage of 'economic affairs and services' and 'education', minus 'social security and welfare', in total public expenditure	IMF
Social consumption	Percentage of 'sickness and maternity', 'unemployment', and 'industrial injuries', minus 'pension', in total 'social security and welfare'	ILO
Labour union movement	Percentage of employees who participate in unions	OECD
Economic modernization	Percentage of non-farmers in all of the labour force	OECD
Non-coverage of pensions	Percentage of aged persons who are not pensioners	WB
Gender discrimination	Gender wage lag	ILO
Stratification in welfare	Number of pension and health insurance schemes	OECD
Self-reliance in retired life	Percentage of elderly incomes not from public pension	SSA
Contribution from employees	Percentage of social security contributions coming from employees	ILO
Family supports	Rate of aged persons living with their children	ECHP
Contribution from employers	Percentage of social security contributions coming from employers	ILO
Scale of private pensions	Scale of private pensions as a percentage of total pensions	OECD
Dependency on trade	Trade ratio to GDP	OECD
Resource dependency	The percentage of non-farm production in GDP	OECD

Principais Etapas

- 1. Escolha das Variáveis**
- 2. Escolha da Medida de Distância**
- 3. Escolha do Algoritmo de Agregação**
- 4. Escolha do Número de Clusters**

Principais Etapas

1. Escolha das Variáveis

2. Escolha da Medida de Distância

3. Escolha do Algoritmo de Agregação

4. Escolha do Número de Clusters

- A seleção das variáveis a incluir pode ser definida por referência a um referencial teórico, ou pode ser um processo indutivo/iterativo.
- A análise de cluster hierárquica pode lidar com dados nominais, ordinais e de escala; no entanto, não é recomendado misturar diferentes níveis de medição.
- Se as variáveis tiverem escalas e médias diferentes, podemos querer padronizar as pontuações Z ou centralizar a escala.

Principais Etapas

1. Escolha das Variáveis

2. Escolha da Medida de Distância

3. Escolha do Algoritmo de Agregação

4. Escolha do Número de Clusters

➤ A escolha da medida de distância depende da escala de medição das variáveis.

➤ Variáveis Contínuas => Distância Euclidiana.

➤ Variáveis Ordinais => Qui-Quadrado

➤ Variáveis Nominais (binárias) => Distância Euclidiana.

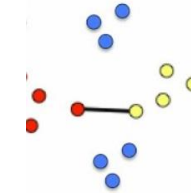
➤ Variáveis Intervalares (binárias) => Distância Euclidiana Quadrada.

Principais Etapas

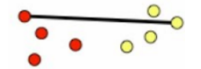
1. Escolha das Variáveis
2. Escolha da Medida de Distância
3. **Escolha do Algoritmo de Agregação**
4. Escolha do Número de Clusters

➤ Temos uma multiplicidade de algoritmos à nossa escolha:

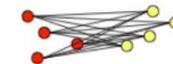
Single Linkage /
Nearest Neighbor



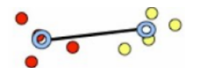
Complete Linkage /
Furthest Neighbor



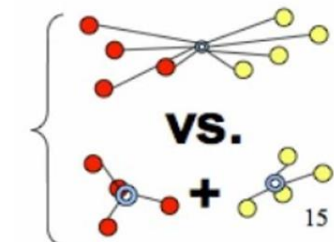
Average Linkage /
Between-Groups



Centroids



Ward's Method



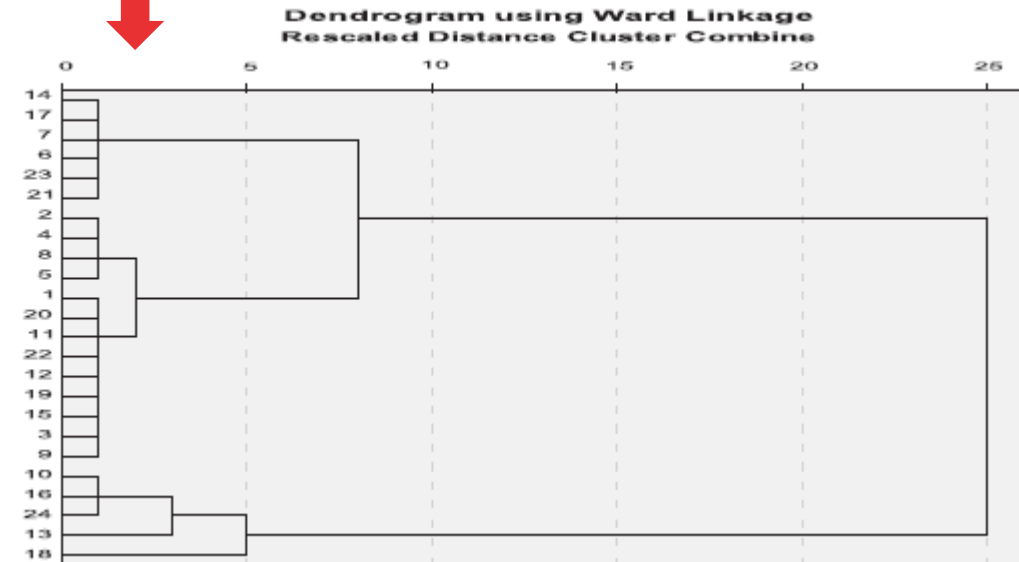
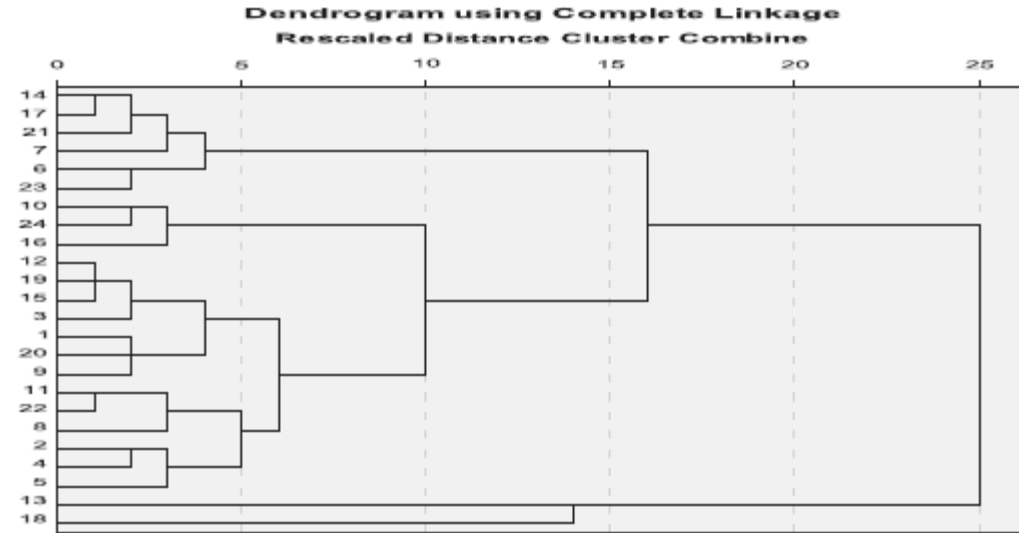
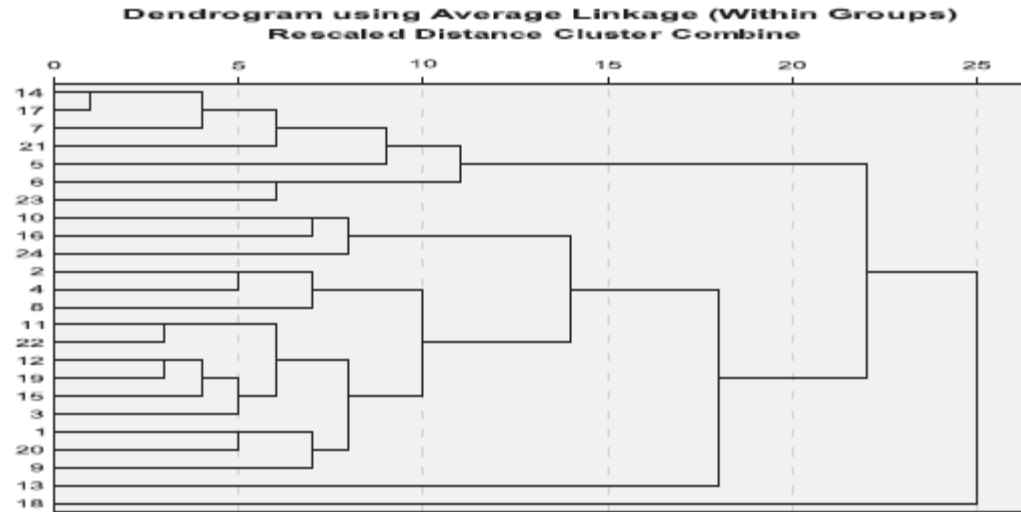
Principais Etapas

1. Escolha das Variáveis
2. Escolha da Medida de Distância
3. **Escolha do Algoritmo de Agregação**
4. Escolha do Número de Clusters

➤ A escolha do Algoritmo de Agregação é particularmente relevante uma vez que, dependendo do algoritmo escolhido o número e composição dos clusters irá variar.

Métodos e Técnicas de Investigação

Aula 4: Descrever a Diversidade na Realidade Social



Principais Etapas

1. Escolha das Variáveis
2. Escolha da Medida de Distância
3. Escolha do Algoritmo de Agregação
4. Escolha do Número de Clusters

➤ Em termos estatísticos, há um número ideal de clusters, que é definido por referência ao 'Coeficiente de Aglomeração'

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	6	.026	0	0	8
2	2	5	.078	0	0	7
3	4	9	.224	0	0	5
4	1	7	.409	0	0	6
5	4	10	.849	3	0	8
6	1	8	1.456	4	0	7
7	1	2	4.503	6	2	9
8	3	4	9.878	1	5	9
9	1	3	18.000	7	8	0

Source: <https://www.slideshare.net/animesh7392/spss-tutorialclusteranalysis>

Análise de Clusters

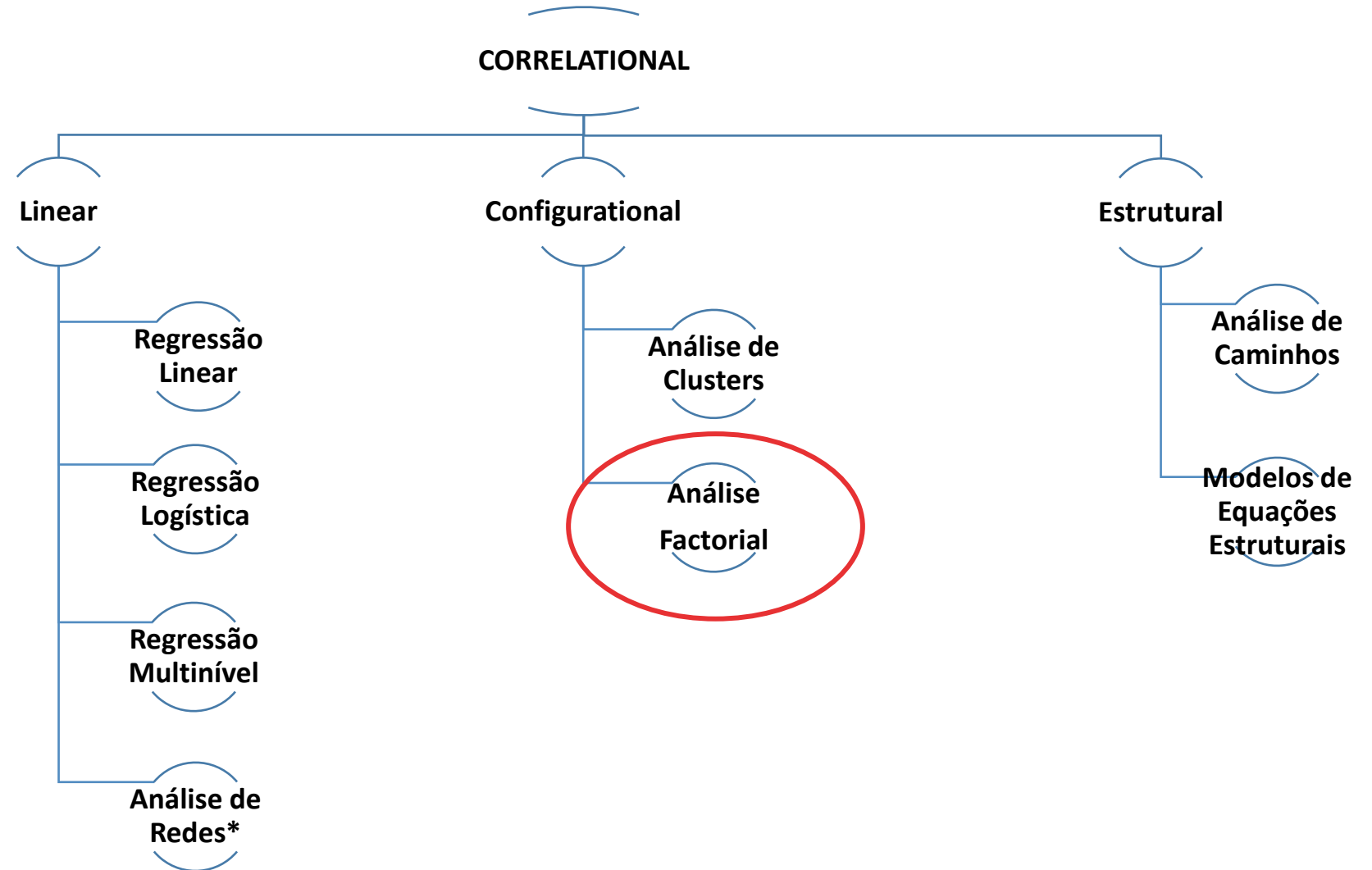
- Origem na Antropologia
- Especialmente útil para identificar a existência de grupos de observações (indivíduos/empresas/países) que partilham entre si um conjunto de características.
- **Pode ser utilizada em combinação o aos modelos de regressão, seja para determinar a variável dependente ou para definir a variável independente.**



Métodos e Técnicas de Investigação

Aula 4: Descrever a Diversidade na Realidade Social

- As you can see, there is large variety of Correlational Research Designs...



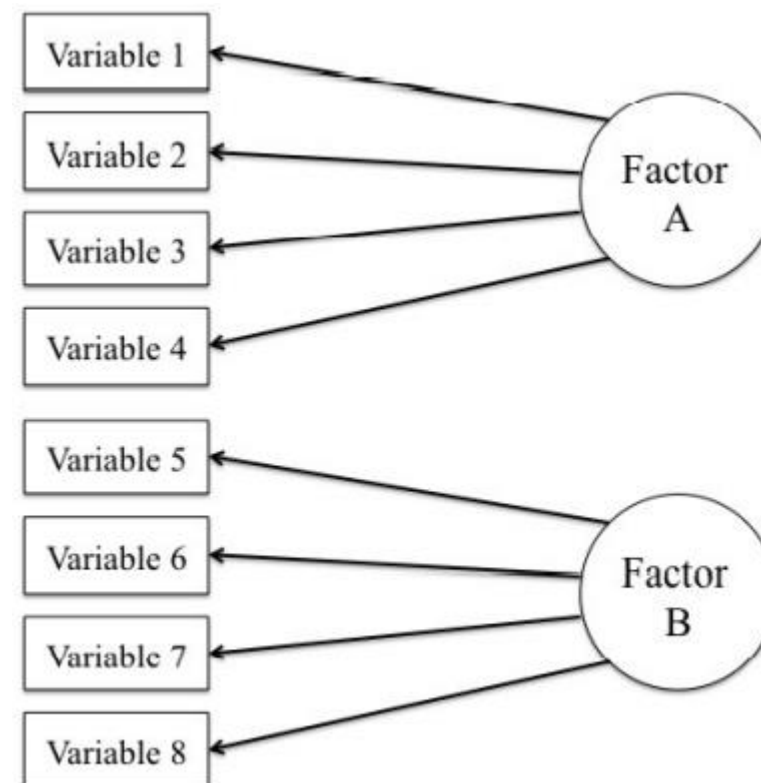
Análise Fatorial (Exploratória)

- A análise fatorial tem suas origens no início de 1900, com o interesse de Charles Spearman na capacidade humana e é amplamente utilizada na psicologia social e em outras ciências sociais.
- A análise fatorial foi originalmente um método estatístico exploratório, posteriormente passou a ser utilizado o teste de hipóteses - análise fatorial confirmatória.



WHY/WHEN USE EXPLORATORY FACTOR ANALYSIS?

- Um fator é um construto ou dimensão que pode explicar as relações (correlações) entre variáveis.
- A análise fatorial exploratória é ideal quando os dados são complexos e não há certeza de quais são as variáveis mais importantes no campo.
- Também pode ser usado para avaliar a validade de escalas.
- A análise fatorial também pode ser usada para desenvolver teoria sobre a natureza do construto.



Principais Etapas

1. Escolha das Variáveis e da Amostra

2. Escolha da Número de Fatores

3. Escolha do Método de Rotação

4. Nomear/Classificar Fatores

- A seleção das variáveis a incluir pode ser definida por referência a um referencial teórico, ou pode ser um processo indutivo/iterativo.
- Para que a AF produza soluções estáveis, a seleção do tamanho da amostra e do número de variáveis deve seguir uma série de regras práticas:
 - A amostra deverá ter, no mínimo, entre 200 e 300 observações;
 - A razão entre o número de observações e o número de variáveis deverá ser, no mínimo, 2:1.

Principais Etapas

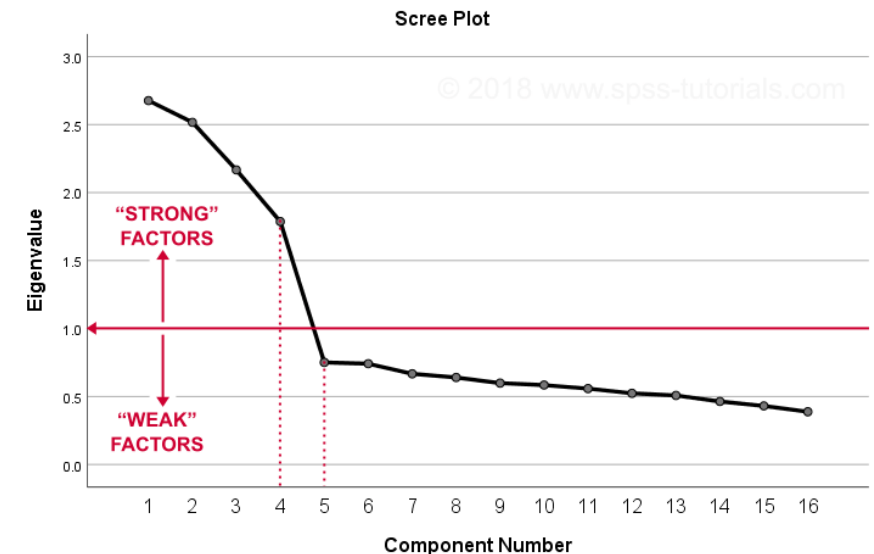
1. Escolha das Variáveis e da Amostra
2. Escolha da Número de Fatores
3. Escolha do Método de Rotação
4. Nomear/Classificar Fatores

➤ Este passo consiste em escolher o tipo de número ideal de fatores que estão subjacentes ao conjunto de variáveis que estão a ser analisadas.

Principais Etapas

1. Escolha das Variáveis e da Amostra
2. Escolha da Número de Fatores
3. Escolha do Método de Rotação
4. Nomear/Classificar Fatores

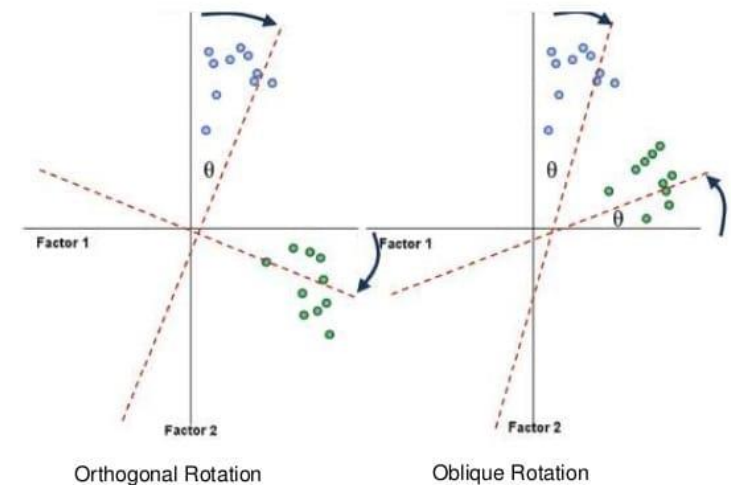
- Esta decisão baseia-se numa medida, o 'Eigenvalue', que é normalmente produzido por pacotes estatísticos;
- Normalmente, devem ser incluídos fatores que possuem um 'Eigenvalue' superior a 1 – regra Kaiser 1 (K1).



Principais Etapas

1. Escolha das Variáveis e da Amostra
2. Escolha da Número de Fatores
3. Escolha do Método de Rotação
4. Nomear/Classificar Fatores

➤ Ao contrário da Análise de Clusters, a Análise Factorial Exploratória é menos sensível à escolha do Método de Rotação. Por norma usa-se o Método de Rotação Ortogonal.



Principais Etapas

1. Escolha das Variáveis e da Amostra

2. Escolha da Número de Fatores

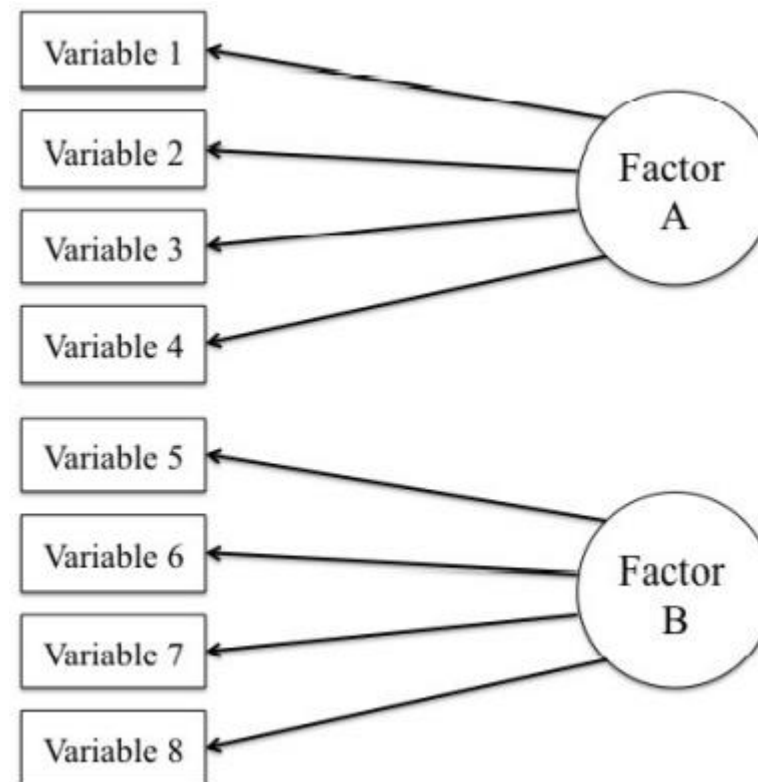
3. Escolha do Método de Rotação

4. **Nomear/Classificar Fatores**

- O passo final é rotular cada fator ou componente de forma a representar como um todo o significado conceitual de cada variável que define uma determinada dimensão latente.
- Caso contrário, podemos usar cargas fatoriais para identificar a natureza do fator.
- As cargas fatoriais indicam a importância relativa de cada variável para cada fator. As variáveis com as cargas mais fortes em um fator específico podem ser usadas para “definir” o fator.

LIMITAÇÕES DA ANÁLISE FATORIAL (EXPLORATÓRIA)

- Existe uma infinidade de soluções matematicamente equivalentes.
- É difícil replicar.
- *'You only get what you put in'* – a exclusão de variáveis pode distorcer os resultados.



Por hoje é tudo...

Até à próxima semana!