DATA SCIENCE & ANALYTICS

MBA

ANÁLISE FATORIAL

F6206851

0000011

Table of Contents

1. INTRODUÇÃO

A Análise Fatorial é uma técnica estatística utilizada para identificar a estrutura subjacente de um conjunto de variáveis observadas. O objetivo principal é reduzir a complexidade dos dados, identificando um número menor de fatores latentes que explicam as correlações entre as variáveis. Esses fatores latentes são construções teóricas que não são diretamente observáveis, mas podem ser inferidas a partir das variáveis medidas.

Existem diferentes métodos de análise fatorial, sendo os mais comuns a Análise Fatorial Exploratória (EFA) e a Análise Fatorial Confirmatória (CFA). A EFA é utilizada para descobrir a estrutura subjacente sem impor uma estrutura preconcebida, enquanto a CFA é utilizada para testar hipóteses específicas sobre a estrutura dos fatores.

1.1. METODOLOGIA DA ANÁLISE FATORIAL

- Seleção das Variáveis:

A primeira etapa é a seleção de variáveis que se acredita estarem correlacionadas e que podem ser explicadas por fatores latentes.

- Matriz de Correlações:

Calcula-se a matriz de correlações entre as variáveis selecionadas para verificar se há correlações suficientes que justifiquem a aplicação da análise fatorial.

- Extração de Fatores:

Utilizam-se métodos como a Análise de Componentes Principais (PCA) ou a Análise Fatorial Comum para extrair os fatores. Na PCA, os fatores extraídos são combinações lineares das variáveis originais que explicam a maior parte da variância.

- Rotação de Fatores:

Após a extração, os fatores podem ser rotacionados para uma interpretação mais clara. Métodos de rotação comuns incluem a rotação varimax (ortogonal) e a rotação oblimin (oblíqua).

- Interpretação dos Fatores:

Os fatores são interpretados com base nas cargas fatoriais, que representam as correlações entre as variáveis originais e os fatores extraídos. Fatores com cargas elevadas em um conjunto de variáveis podem ser interpretados como representando uma dimensão latente específica.

1.2. Vantagens e Limitações

- Vantagens:

  - Redução da complexidade dos dados.
  - Identificação de estruturas subjacentes.
  - Facilitação da interpretação de grandes conjuntos de dados.

- Limitações:

  - Requer grandes amostras para resultados confiáveis.
  - A interpretação dos fatores pode ser subjetiva.
  - A escolha do número de fatores a ser retido pode ser arbitrária.

1.3. Aplicações da Análise Fatorial

A Análise Fatorial tem uma ampla gama de aplicações em diversas disciplinas, incluindo:

- Psicologia:

  - Identificação de dimensões subjacentes a traços de personalidade, atitudes e comportamentos.
  - Desenvolvimento e validação de escalas de medição psicológica.

- Educação:

  - Análise de dados de testes para identificar habilidades ou competências subjacentes.
  - Desenvolvimento de instrumentos de avaliação educacional.

- Marketing:

    - Segmentação de mercado baseada em características de consumidores.
    - Identificação de fatores que influenciam a percepção e a satisfação do cliente.

- Ciências Sociais:

    - Exploração de estruturas latentes em pesquisas sociais, como valores culturais e atitudes políticas.
    - Análise de dados de pesquisas para identificar padrões de comportamento social.

- Economia e Finanças:

    - Redução da dimensionalidade de conjuntos de dados financeiros para identificar fatores de risco.
    - Análise de dados econômicos para identificar padrões subjacentes de crescimento e desenvolvimento.

2. CONTEXTUALIZAÇÃO

A análise fatorial é aplicada quando se trabalha com variáveis métricas, dependendo das correlações entre essas variáveis. O principal objetivo é agrupar variáveis em fatores para entender seu comportamento conjunto, reduzir a estrutura dos dados, validar construtos, criar rankings de desempenho e gerar fatores ortogonais para uso em modelos supervisionados.

2.1. QUANDO APLICAR A ANÁLISE FATORIAL?

- Quando as variáveis forem métricas: depende das correlações entre variáveis;
- Trata-se do agrupamento das variáveis em fatores. Os objetivos podem ser:
    - Obter o comportamento conjunto de variáveis, combinando-as para redução estrutural;
    - Análise da validade de construtos pela identificação das variáveis alocadas aos fatores;
    - Elaboração de rankings para classificação de desempenho por meio dos fatores;
    - Criação de fatores ortogonais entre eles e posterior uso em modelos supervisionados.

- Análise fatorial por componentes principais:

A Análise Fatorial por Componentes Principais (PCA) é uma técnica estatística utilizada para reduzir a dimensionalidade dos dados, transformando variáveis correlacionadas em um conjunto menor de fatores não correlacionados entre si. Esses fatores são extraídos com base nos autovalores, e o critério de Kaiser recomenda a retenção de fatores com autovalores maiores que 1.

- método de determinação dos fatores que se baseia na criação de fatores não correlacionados a partir da combinação linear das variáveis originais;
- modelo não supervisionado de machine learning - portanto, a técnica não tem um caráter preditivo para observações que

não estejam presentes na amostra. Se surgirem novas observações, novos fatores atualizados devem ser gerados;

- A análise fatorial é aplicada quando se trabalha com variáveis métricas, dependendo das correlações entre essas variáveis.

- O principal objetivo é agrupar variáveis em fatores para entender seu comportamento conjunto, reduzir a estrutura dos dados, validar construtos, criar rankings de desempenho e gerar fatores ortogonais para uso em modelos supervisionados.

3. IMPLEMENTAÇÃO

Para aplicar a PCA, as variáveis devem ser métricas e correlacionadas.

1) A análise começa com a matriz de correlações de Pearson,
2) e a adequação da análise é verificada pelo teste de esfericidade de Bartlett.
3) Na criação de rankings a partir dos fatores extraídos, utiliza-se a soma ponderada dos fatores, onde cada fator é ponderado pelo seu percentual de variância compartilhada.
4) Scores fatoriais são valores quantitativos que indicam a posição das observações em relação aos fatores.

3.1. Matriz de correlações de Pearson

A implementação da PCA começa com a matriz de correlações de Pearson, que avalia a relação linear entre variáveis métricas. Coeficientes próximos de -1 ou +1 indicam correlações fortes, enquanto coeficientes próximos de zero indicam baixa correlação.

Se as correlações entre variáveis forem próximas a zero e sem significância estatística, a análise fatorial não é adequada, pois não há correlações significativas.

- Coeficiente de correlação de Pearson: relação linear entre duas variáveis métricas;

- Coeficientes de correlação mais próximos dos valores extremos (-1; +1) propiciam a extração de um único fator → indicam existência de relação entre as variáveis;

- Coeficientes de correlação mais próximos de zero propiciam a extração de diferentes fatores → indicam que a relação entre as variáveis é (praticamente) inexistente.

3.2. ADEQUAÇÃO GLOBAL DA ANÁLISE

A adequação da análise é verificada pelo teste de esfericidade de Bartlett, que compara a matriz de correlações com a matriz identidade; um p-valor menor que 0.05 indica que a matriz de correlações é significativamente diferente da matriz identidade, justificando a aplicação da análise fatorial.

- Para que que a análise fatorial seja adequada, devem existir valores mais elevados (-1; +1) e estatisticamente significantes na matriz de correlações;

- Para investigar a adequação global da análise fatorial, vamos utilizar o teste de esfericidade de Bartlett;

- Os coeficientes de correlação de Pearson são estatisticamente diferentes de zero?

- Teste de esfericidade de Bartlett: Compara a matriz de correlações com a matriz identidade de mesma dimensão e espera-se que tais matrizes sejam diferentes para que a análise seja aplicável

3.3. AUTOVALORES E AUTOVETORES

Os autovalores indicam o percentual da variância explicada por cada fator, enquanto os autovetores são derivados desses autovalores.

Esses elementos são essenciais para identificar os scores fatoriais, que relacionam os fatores com as variáveis originais em um modelo linear. Para K variáveis originais, podem existir no máximo K fatores, com os scores sendo obtidos dos autovalores e autovetores.

- Os autovalores indicam o percentual da variância compartilhada pelas variáveis originais para a formação de cada fator;
- Os autovetores da matriz de correlações são obtidos com base em cada um dos autovalores

3.4. Obtenção dos fatores

Nem todos os K fatores serão necessariamente utilizados. A análise dos autovalores é fundamental, onde somente fatores com autovalores maiores que 1 (critério de Kaiser) são considerados representativos.

As cargas fatoriais representam a correlação entre os fatores e as variáveis originais, indicando a importância de cada variável na constituição dos fatores, indicando que valores elevados de carga fatorial refletem alta correlação.

Em uma PCA, um score fatorial igual a zero indica que a observação correspondente não tem influência relevante no fator em análise.

Em relação aos fatores extraídos na análise fatorial, os valores calculados para as observações da amostra com base nos fatores formam variáveis quantitativas. Esses valores, conhecidos como scores fatoriais, são medidas contínuas que representam a posição de cada observação em relação aos fatores extraídos.

- Identificação dos scores fatoriais;
- Após a análise fatorial ser considerada adequada pelos testes anteriores, será necessário criar os scores que geram os fatores propriamente ditos;
- Scores fatoriais: são os parâmetros que relacionam o fator com as variáveis originais, representados em um modelo linear;
- Para K variáveis originais, existem, no máximo, K fatores (F1, F2, ... , Fk);
- Os scores vêm a partir dos autovalores e autovetores da matriz de correlações;

3.5. Escolha dos fatores

▪ Todos os K fatores serão utilizados?

▪ Embora seja possível estabelecer a priori quantos fatores são desejados, é de fundamental importância realizar uma análise por meio dos autovalores

▪ Lembrando: os autovalores indicam o percentual da variância compartilhada pelas variáveis originais para a formação de cada fator

▪ Neste sentido, fatores formados a partir de autovalores menores do que 1 podem não ter representatividade. O critério de Kaiser (ou critério da raiz latente) indica que sejam considerados apenas fatores correspondentes a autovalores > 1

3.6. Cargas fatoriais

▪ Análise da composição dos fatores

▪ As cargas fatoriais representam as correlações de Pearson entre os fatores e as variáveis originais

▪ Pode ser interpretada como a importância de cada variável na constituição daquele fator em particular

▪ Quanto maior a carga fatorial, mais aquele fator é influenciado pela variável

3.7. Comunalidades

Comunalidades mostram a variância total compartilhada por cada variável nos fatores selecionados, permitindo analisar a perda de variância após a exclusão de fatores não significativos, sendo que valores baixos indicam maior perda de variância.

- Composição dos fatores selecionados;
- Ao utilizar o critério da raiz latente, somente os fatores que são derivados de autovalores maiores que 1 serão considerados;
- Portanto, as comunalidades mostram a variância total compartilhada, para cada variável, em todos os fatores extraídos e selecionados com base no critério da raiz latente;

- É possível analisar se houve perda de variância, por variável, após a exclusão de fatores por meio do critério da raiz latente;

3.8. Criação de rankings

Para criar rankings, utiliza-se a soma ponderada e ordenamento dos fatores. Cada observação na amostra tem seu resultado calculado multiplicando-se o valor de cada fator pelo seu percentual de variância explicada, e depois ordenando esses resultados para formar o ranking.

- Soma ponderada e ordenamento;

- Para criar rankings a partir dos fatores obtidos utilizando o critério da soma ponderada e ordenamento, para cada observação da amostra, calcula-se:
- Resultadoi = (F1i * % var. comp. F1) + (F2i * % var. comp. F2) + ... + (Fki * % var. comp. Fk)
- Em resumo, multiplica-se o resultado obtido de cada fator por seu percentual de variância compartilhada e depois é realizado o ordenamento do resultado